DOCTOR OF PHILOSOPHY

Population genetic markers in biomonitoring programmes : a case study of flatfish around the British Isles

Tysklind, Niklas

Award date:
2009

Awarding institution:
Bangor University

PRIFYSGOL BANGOR UNIVERSITY

Cefas

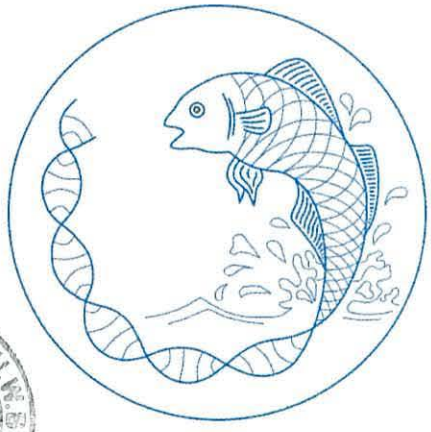
Population Genetic Markers in Biomonitoring Programmes: A Case Study of Flatfish Around the British Isles

A thesis submitted to Bangor University, Wales, for the Degree of Doctor of Philosophy

By Niklas Tysklind, B.Sc.

August 2009

Molecular Ecology and Fisheries Genetics Laboratory
Environment Centre for Wales, 3rd floor
School of Biological Sciences
Bangor University
Deniol Road, LL57 2UW



To Nina, my dear aunt and second mum...

may you always be with me

Siento que la vida se me escapa entre mis garras,
como arena seca e inerte incapaz de construir castillos,
al ver como los que estuvieron antes que yo,
que explotaron al máximo su existencia,
sus locomotoras rugiendo con la fuerza de la tormenta,
felizmente se arrodillan ante lo inevitable,
y aceptan ver como se apaga la vela de sus vidas.

Siento un inconmensurable vacio interior,
al descubrir que todo el saber que de ellos ingerí,
no es más que la punta del iceberg de su experiencia,
que las locuras, los descubrimientos, y los desgarros del alma,
que debieron formar los intrincados castillos de cartas de sus vidas,
se hallan ahora fuera del alcance de mis redes.

Siento el ir y venir de la marea de emociones,
contundiendo los muelles de mi más protegido puerto,
cada vez que recuerdo los momentos vividos junto a ellos,
las risas, las historias, las charlas hasta las tantas, las reyertas,
pero sobre todo, la compañía de sus mentes,
ahora en un estado incomprensiblemente etéreo.

Siento un terror más fuerte que a la misma muerte,
al pensar en la posibilidad de la desintegración de lo conocido,
que la efimeridad de lo corporal pueda infectarse al alma,
ligándola a un destino aun más perecedero que la recombinación de lo material,
y amenazando con rajar las infladas velas de mi orgulloso pero delicado galeón.

Siento la más pura serenidad al saber que navegan junto a mí,
guiando mi proa y sondeando el camino por trazar,
hasta el día el que sean mis propios engranajes los que dejen de gruñir,
y me una a ellos en un eterno gozo del saber infinito,
guiando y velando por aquellos que empiezan su cruzada por el océano.

Abstract

The use of bioindicator fish species in monitoring programmes is widely accepted as a means of assessing marine environmental health. In such approaches, the effects of pollutants and other anthropogenic impacts on the health of individual fish is evaluated and compared to that of non-exposed reference fish. However, not all fish are the same. Genetic diversity between individuals translates into variance on how and when individuals respond to pollution, and the combined response of interbreeding individuals is reflected on how a population is affected by exposure to environmental stressors such as pollutants.

Whether due to natural causes or the result of previous pollution exposure and selection, some populations may be more tolerant to pollutant exposure than others. If not corrected, comparisons of disease profiles among populations with differing pollution tolerance levels will affect the interpretation of results from biomonitoring programmes.

In the UK, dab, *Limanda limanda,* and flounder, *Platichthys flesus,* are routinely used as environmental bioindicators of pollutant exposure; however, little information exists on population structuring among sampling locations. Here, the development of neutral microsatellite markers for dab and flounder is described. A novel approach for reducing the cost of labelling microsatellite primers in combination with multiple amplification of several loci in a single tube is then devised. Next, the development and preliminary evaluation of adaptive genetic markers to detect selection imposed by pollution is reported. No definitive evidence of strong and recent selective pressures at the analyzed genes is found, but suggestions for future research are made. Estimating genetic differentiation between populations is central in population genetic studies. Several new and traditional estimators of genetic differentiation are compared empirically. Consecutively, the genetic structure of dab around the British Isles is analysed and described. Two main dab subpopulations, subtly but significantly differentiated, have been identified, corresponding to the North Sea and Irish Sea basins, though there is also evidence of structuring at other scales. The implications for biomonitoring programmes are considered. Finally the combination of both genetic and biomonitoring information is explored. No evidence of increased relatedness or inbreeding among individuals afflicted with liver nodules is found, however, for some samples with abnormally high frequency of liver nodules, the incidence can be explained by recent immigration from other locations. Genotyping of assessed individuals provided important information not available by other means and the incorporation of population genetic data is encouraged for biomonitoring programmes studying mobile species.

Acknowledgements

I would like to start thanking my superb and diverse supervisory board, who have all excelled at providing guidance and advice on the topics covered in this thesis. The multifaceted nature of which demanded a wide range of expertise, and you have all wilfully provided the best advice possible.

I would like to express my most sincere gratitude to Prof. Gary Carvalho for his continuous encouragement, and broad range of advice received during the last four years; for all those moments when I lost faith and needed reassurance; for knowing how to bend the rules when needed; for all those conferences he's managed to send me to; for surviving his "dead brain" after reading previous drafts of this thesis, and still be able to provide constructive comments; for all the dinners and wines and laughter; but most of all for believing in me.

To Dr. Brett Lyons, for all the help with sampling collection; for all the guidance regarding biomonitoring points of views; for providing all the fish disease records used across the thesis; for finding ways of keeping me financially alive during the last few months; and for those unforgettable sampling cruises around the British Isles disembowelling unsuspecting fish.

To Dr. Martin Taylor, for all that invaluable advice on microsatellite construction; for providing a keen and critical eye on statistical issues, without which my progress would have been halved; for raising the standards and bringing me back to Earth when I was in "Blake 7"; for introducing me to the underworld of R; for those moments of recognition that made me feel like the king of world (or the underworld?); and for helping me break through my creativity walls.

To Dr. Ian McCarthy, who contributed with his extensive knowledge of fish biology and endured all those long meetings and manuscripts of unintelligible "genetic non-sense"; thanks for all your styling and wording advice; for pointing out what is only obvious to geneticists (we often forget there are people out there who do not do PCRs for a living); and for providing the right encouragement when it was most needed.

I would like to thank both Bangor University and CEFAS for the joint studentship that allowed me undertake this thesis, and to DEFRA who have funded the project: Grant N°. ME3206.

There is a long list of people who have contributed in one way or another towards making this thesis possible and turning these last four years into such an enriching personal experience.

Thanks to Dr. Teresa Neuparth and Gregg Ashcroft who helped with the flounder microsatellite library isolation and subsequent testing of the microsatellites. A very special thanks goes to Freya Goodsir, who spent long hours doing DNA extractions. I am thankful to Yann Surget, Sonia Pascoal, and Serinde van Wick for trialling the oligo-tails and reporting their results with anoles, snails and guppies. I am indebted to Dr. Hans Gerritsen from the Marine Institute, Galway (EI) for kindly supplying the Atlantic dab samples. Thanks to all the CEFAS *Endeavour* scientific crew, who collected all the biomonitoring data, and who are now out there sampling for more dab; thanks for the company while at sea, and for all the laughter playing *Buzz*; for the beers in Weymouth; and the salsa on board.

Thanks to Professor Paul Dando for sharing his wealth of knowledge on flounder biology; Thanks to Dr. Lou Jost and Nick Crawford for assistance and discussion on calculating genetic distances. A huge thank you must go to Dr. Lorenz Hauser, Dr. Sarah Helyar, Dr. Jenny Rock, and Dr. Yann Surget for so much interest, and many discussions and comments about estimators of differentiation, relatedness, and migration. Really enjoyed those discussions! Thanks to Dr. Thomas Caspari for hints on human cancer incidence and research. I would like to thank Dr. Thibaut Jombart for invaluable help with the multivariate analysis. The production of the map graphics would have been impossible without the help of Prof. Roger Bivand. Thanks to Yves from Codeplex.com, who provided much needed advice about how to break into Microsoft Word and create new referencing styles. Niklas 1: Microsoft 0!!

Thanks to Marine Genomics Europe and all the people involved (both lecturers and students) for all the courses I was allowed to attend and where I learned so much about analysis of the genome at so many levels, and for critical thinking and discussion on genetic issues. These courses made the last four years much more exciting. Thanks in particular to Dr. Anti Vasemägi and Dr. Galice Hoarau for helpful discussions on gene-linked microsatellites, and thanks to Dr. Jeanine Olsen from whom I learned that sometimes humility is not a priced quality in science.

I would like to thank all the fantastic people in the Molecol lab: Cathy, Pete, Polly, Si, Sonia, Yann, Karen, Sam, Jacky, Silvain, Helena, Filipe, Alix, Nick & Lou, Marko, Delphine, Aitor, Serinde, Andy, Vera, Axel, and Munny for all the help; the stimulating conversations; the fun;

the music wars; the visits of the "president"; the Friday night pints and all those dinners that made this lab so special. You are the best!

My gratitude goes to Wendy, our excellent lab manager, for all the help during the last four years. A special mention goes to Jan, for all the laughter and innumerable amount of times she kindly offered her help with one thing or another, and for so much help while desperately printing this thesis. Thanks to Jane, Gillian, Karen, Sally, Sim, and Emlyn for your patience and all those last minute little things you helped me with. Thanks everyone else working in the Brambell, Memorial and ECW for all the help received in the last four years; for the healthy distractions and interesting conversations; you're all such a great gang.

Thanks to "I hate fish" Gavan and "venom" Nick, who coped with my expansive mess in the office during the last few months. Thanks to Jenny for all the moral support and making me feel human in the long hours in the bioinformatics room, I'll never forget you words of wisdom: "Fish are weird" and "I don't believe in microsatellites". Like Mastercard... Priceless! Thanks to Joana, my conference pal, for all the company, the dances, the moans and the coffees together.

My experience as a postgraduate student would not have been as complete without my involvement with the Bangor University Research Students Forum committee, where I learned so much about teams, people, organising conferences and "Cheese & Wines". Thanks to Dr. Sue Assinder and Dr. Penny Dowdney for giving me the opportunity to be the post-grad student-rep at Change Academy; that was an unforgettable experience.

Thanks to all the people at Salsa Bangor whose company, dancing skills, and moral support on Thursday nights kept me alive and kicking when not much else made me move.

A very special thought goes to the long list of different housemates who have passed through "53 Caellepa", and made my life so much richer: my brother Alek, Esther, Itxaso, Vicky, Julia, Evi, Becky, Lindsay, Marek, Luis, Jesse, Jon... And all the other "53 Caellepa add-ons" that always made living there so special: Spanish Bea and the other Spanish Bea, Phil, Max, Walter, Brian, all my fellow lab-mates, the list goes on... who have been my "family" here and endured my mood changes and accepted my absences cheerfully.

Thanks to Vicky for all the DIYing help; for the discussions on genetics; for the company; and for reading through all those chapter and providing comments on them, I owe you one!

Thanks to Lindsay, my endurance partner in this long seemingly never-ending marathon; for all her company and moral support; for those long evenings sipping wine that seem like decades ago (we will have some soon); for the tears; and for being the first of our generation opening the path for the rest of us to come; for being there... I do not know how I would have survived this without you. Best of lucks tomorrow!!!

I would like to thank Gregg for coping with my bad temper and anger; for cheering me up; and for always being there, no matter what, every time I needed someone.

I thank the Sun, for rising every morning across the office and the dormant city, giving me a glimmer of hope and perspective in those darkest hours of desperation.

And finally I would like to thank my parents and family, for their unconditional faith in me; for coping with my absences; for listening to me and guiding me when I lost my way; for lifting my spirits up when I felt down; and for their unlimited love. *Sois los mejores!!!*

List of Publications

Tysklind N, Neuparth T, Ashcroft GR, Taylor MI, Lyons BP, McCarthy ID, Carvalho GR (2009) Isolation and characterization of 28 new microsatellite markers for European flounder (*Platichthys flesus* L.), *Molecular Ecology Resources*, 9, 1065-1068.

Tysklind N, Taylor MI, Lyons BP, McCarthy ID, Carvalho GR (2009) Development of 30 microsatellite markers for dab (*Limanda limanda* L.): a key UK biomonitoring species, *Molecular Ecology Resources*, 9, 951-955.

Index of Contents

Index of Tables

Index of Figures

Glossary of Terms:

Adenocarcinoma: a malignant cancer originating from a glandular tissue.

Adenoma: a benign tumour in glandular tissue.

Allele frequency: the relative proportion of individuals carrying a particular allele at a certain locus as estimated from a population sample.

Allele: a variant in a polymorphic loci.

Allozyme: polymorphic enzyme varying in their amino acid sequence which can be used as a genetic marker.

Amplicon: the target of a PCR reaction, which is copied multiple times.

Amplify/amplification: the process by which a target DNA sequence (amplicon) is copied many times in a PCR reaction.

Bioindicators: a focal species used in biomonitoring for the assessment of pollution impacts on the natural environment.

Biomarker: a measurable aspect of the biology of a certain species modified by the presence of pollutants.

Biomonitoring: the assessment of the ecosystem health my measurement of species diversity or health status of a particular species.

bp (base-pair): the two complementary DNA bases on either DNA strand, which can be used as unit to measure the length of a DNA sequence.

Cancer: a disease, in which a cell engages in uncontrolled growth, invasion of nearby tissues, and possibly metastasis.

Carcinogen: a substance believed to induce cancerigenous processes.

Carcinoma: a malignant cancer originating from epithelial tissue.

EST (Expressed sequence tag): a DNA sequence partially representing a transcribed gene.

Genetic drift: the fortuitous changes in reproductive output between individuals which lead to changes in allele frequencies between generations.

Genome: the full hereditary information of an organism encoded by DNA.

H_E: Expected frequency of heterozygotes at a locus under HWE given the allelic diversity at the locus. Due to its tight relationship with number of alleles it is also known as gene diversity.

Heavy metal: a loose term referring to a subset of chemical elements with heavy atomic weight.

Hepatocarcinoma: a malignant cancer originating in the liver.

Heterozygote: an individual with two different alleles for the same locus

H_O: Observed frequency of heterozygotes at a given locus.

Homoplasy: within a locus, the mutation of an allele to another already pre-existing allelic state, thus not resulting in a new detectable allele.

Homozygote: an individual with two equal alleles for the same locus.

H_S: Within-sample Heterozygosity: Expected heterozygosity (H_E) within a single sample.

H_T: Total heterozygosity: Expected heterozygosity (H_E) calculated from a combination of samples.

HWE (Hardy-Weinberg Equilibrium): The Hardy-Weinberg model states that, if two alleles of the same locus, A and B, are present in frequencies p and q (where $p + q = 1$), then the frequencies of homozygotes for A and B are AA = p^2 and BB = q^2, while heterozygotes will be present at AB = $2pq$; (note that: $p^2 + q^2 + 2pq = 1$). In the absence of selection, non-random mating and gene flow, the genotype frequencies remain essentially unchanged across generations. The model can be extended for loci with more than two alleles following the same principle, and larger number of alleles in a locus results in higher expected frequency of heterozygotes under the HWE.

Linkage disequilibrium: the non-random association of alleles at two different loci.

Lipophilic: ability of a substance to dissolve in non-polar substances such as oils, fats and other lipids.

Locus/loci: a certain location within the genome of an organism.

Microsatellite: a tandemly repeated short DNA sequence, typically present in non-coding parts of the genome, used as a genetic marker.

Migration: from a population genetics perspective, the movement and successful contribution to the next generation of an individual between otherwise independent populations.

Monomorphic: a locus is said to be monomorphic when only one allele is present.

mtDNA: mitochondrial DNA, DNA located within the mitochondria.

Multiplex: the amplification of several different amplicons within a single PCR reaction.

Mutation: the process by which a particular sequence of DNA is changed from one allele to another.

Neoplasms/neoplasia: the abnormal cell growth within a tissue.

Null allele: an undetected allele, either due to mutations in the priming site or scoring errors.

Oncogene: a gene variant which putatively causes cancer.

Oncogenic: the process of genetic and cellular changes leading to the malignant transformation of a cell leading to the formation of a tumour.

PAH: Polycyclic aromatic hydrocarbons.

Panmixia: a state in which migration between all sampling locations is high enough to overcome genetic drift and local selective forces.

PCB: Polychlorinated biphenyls.

PCR (Polymerase chain reaction): a technique employed in molecular biology for the amplification (generating many copies) of a particular DNA sequence (amplicon).

Polymorphic: a locus is said to be polymorphic when more than one allele is present.

Primer: a short nucleic acid of a certain sequence designed to anneal to a particular location in the genome and used as starting point for DNA replication. Two, one forwards and one reverse, are needed in a PCR reaction.

Restriction enzymes: enzymes used to *cut* a DNA sequence at particular places known as restriction sites.

Sample: from a population genetics perspective, a *population sample* is a subset of individuals collected from a location, and thought to be genetically representative of the local population.

SNP: Single nucleotide polymorphism, a particular nucleotide within a DNA sequence for which several alleles are detected among individuals.

SSR: Simple sequence repeat; see microsatellite.

Tumorigenesis: the process of tumour formation.

Chapter 1: Introduction

1.1 Pollution, its Consequences, and Biomonitoring:

1.1.1 Pollution:

Modern human activities have profound impacts on marine environments which can have negative effects on single organisms and entire ecosystems (Hughes *et al.*, 2005). These effects in turn, may interfere with human uses of the natural environment. Overfishing and by-catch are by far the most pressing problems faced by many marine organisms (Pauly *et al.*, 1998; Jackson *et al.*, 2001; Myers & Worm, 2003; Pauly *et al.*, 2005), though, other activities can also impact marine ecosystems at local and global scales. According to Costa *et al.* (2002), non-harvesting anthropogenic impacts on marine ecosystems can be broadly classified into five classes: 1) Changes in temperature regimes: either by localised thermal pollution from power stations (Ahn *et al.*, 2006) or global climate change (Perry *et al.*, 2005); 2) engineering works and extraction of aggregates: which can increase turbidity and sediment deposition (Desprez, 2000); 3) eutrophication of rivers, estuaries and coastal waters which promote noxious red tides and uncontrolled algae growth (Kite-Powell *et al.*, 2008); 4) acoustic pollution created by ship engines, sonar and seismic surveys which mask natural sounds used by cetaceans, fish and other marine life (Weilgart, 2007); and 5) the introduction or rise in concentration of chemical compounds. The current thesis will deal primarily with the effects of the latter.

Exogenous compounds of anthropogenic origin are generally known as pollutants or contaminants, and direct and indirect exposure to pollutants have proven damaging to both humans and other living organisms (Hawkins *et al.*, 1988; Tanabe, 1988; Baumann, 1998; Johnson, 1998; Reichert *et al.*, 1998; Järup, 2003; Waisberg *et al.*, 2003). Once released into the environment, pollutants may move substantially though either transport by air, in water courses or biogenically (Ballschmiter, 1992; Scheringer, 2009), and then concentrate in marine environments such as estuaries, coastal waters and oceans (Haynes & Johnson, 2000; Islam & Tanaka, 2004; Hughes *et al.*, 2005; Kite-Powell *et al.*, 2008). The effects of contaminants on an organism will depend on the nature of the contaminant, whether it is acting singularly or in combination with others, or whether it is interacting with other environmental factors. Furthermore, contaminants may have life stage, sex or population specific effects.

Lawrence & Elliot (2003) classify the most common contaminants in estuaries and coastal areas into three groups: 1) Heavy metals, particularly cadmium, lead, mercury and arsenic, which are bioaccumulated by fish and shellfish and disrupt enzymatic activities, osmoregulatory processes, and the immune system (Pulsford *et al.*, 1995; Johnson, 1998; Järup, 2003); 2) Oils and petrochemicals, which have coating properties in the short term, and are bioaccumulated in the long term and may have tainting effects (change in odour or flavour) on commercial species (Hellou *et al.*, 1995; Hellou & Warren, 1997; Srogi, 2007); 3) Persistent and synthetic organic chemicals, mainly halogenated and non-halogenated hydrocarbons, and organometals (Tanabe, 1988; Marchand *et al.*, 2002). Of all the contaminants, polychlorinated biphenyls (PCBs) and polycyclic aromatic hydrocarbons (PAHs) have received particular attention (Tanabe, 1988; EU Commission, 2001; Baumann, 1998; Kalantzi *et al.*, 2001; Judd *et al.*, 2004; Srogi, 2007). The former are commercially produced by chlorination of biphenyls (Tanabe, 1988); while the latter are the product of incomplete combustion of fossil fuels (Srogi, 2007). These organic chemicals are lipophilic and therefore accumulate in lipid-high tissues within organisms (Tanabe, 1988; Kalantzi *et al.*, 2001), and are considered as direct and indirect cause of cancers (Baumann, 1998; Reichert *et al.*, 1998; Stentiford *et al.*, 2003).

1.1.2 Biomonitoring:

Pollutants can have a dramatic impact on ecosystem health, which here will be defined as the ability of a particular environment to perform ecosystem services beneficial to humans. The consequences of pollution in the services offered by the marine environment include the loss of food items due to poisoning (Judd *et al.*, 2004; Corsolini *et al.*, 2005), reduced fishing yields and loss of leisure areas (Jones, 2006), and the loss of biodiversity (Worm *et al.*, 2006).

On the other hand, reduction or elimination of pollution loads can result in recovery and improved environmental quality. At a localised level, Smith & Shackley (2006) reported a change in benthic fauna towards a more pollution-sensitive species composition in Swansea Bay after the complete cessation of sewage discharge. Myers *et al.* (2008) reported significant improvements of flatfish health in Puget Sound, USA, after sequestering highly polluted deposits underneath a cap of clean sediments. At an ecosystem level, whole estuaries, like the Mersey and Thames, have been the target of intense ecosystem restoration programmes resulting in marked improvements, albeit still incomplete, in biological diversity and ecological rehabilitation (Andrews & Rickard, 1980; Jones, 2006). Therefore, there is interest in understanding the effect of contaminants on the marine environment, assessing the spatial

and temporal trends of those effects, and critically evaluating implications of pollution and the efficiency of contaminant control programmes and policies. That is the aim of biomonitoring: to monitor the biological effects of pollution (Phillips & Segar, 1986).

Biomonitoring programmes assess the health of an ecosystem by either measuring levels of species diversity (Andrews & Rickard, 1980; Whitfield & Elliot, 2002; Breine *et al.*, 2007) or the health status of particular species. In the latter cases, the focal species is known as a *bioindicator*, and some measurable aspect of its biology, a *biomarker*, is modified by the presence of pollutants (Everaarts, 1995). When choosing a bioindicator it is of paramount importance that the species lives in intimate contact with pollutants, that it readily absorbs and processes those pollutants, and that long term effects of pollutant exposure can be measured (Wu *et al.*, 2005). Within the marine environment, species of flatfish fulfil these requirements, and thus they are used in many areas of the world in biomonitoring assessments (MAFF, 1987; Eggens *et al.*, 1995; Hellou & Warren, 1997; Baumann, 1998; Johnson *et al.*, 1998; Nakata *et al.*, 2003; Feist *et al.*, 2004; Davoodi & Claireaux, 2007; Myers *et al.*, 2008; Leonardi *et al.*, 2009).

In the UK, two flatfish species, namely dab and flounder, are of particular importance due to their relatively high abundance and ubiquity in the European coastal shelf. Both species are regularly monitored in the rivers, estuaries and coastal areas of England and Wales, and standard techniques on how to collect and analyse biomarker data (Feist *et al.*, 2004), and accompanying quality assurance procedures (BEQUALM, 2009) have been established. In the next two sections the biological characteristics of the two bioindicator species will be considered, followed by an overview of the biomarkers considered in the biomonitoring programmes.

1.2 Flatfish Biology:

Gibson (2005) provides an extensive review of flatfish biology. Flatfishes comprise a significant component of all teleost fishes, with more than 700 species from over 15 families. All flatfishes share characteristics such as benthic lifestyle, laterally compressed body, having both eyes on the same side of the head, chamaleonic ability to change colour and patterns according to their environment, and the innate skill to bury into the sediment. Eye migration on the larva, a dorsal fin that continues onto the head, and protrusible eyes constitute the characteristic feature of the Order Pleuronectiformes. Several European flatfish species, such as dab, *Limanda limanda,* flounder, *Platichthys flesus,* and Plaice, *Pleuronectes platessa,*

belong to the Pleuronectidae family. Other families represented in European coastal waters are Soleidae (Soles) and Scophthalmidae (Turbot) (Millner *et al.*, 2005).

Verneau *et al.* (1994) composed a phylogenetic tree to understand the evolutionary relationships between Pleuronectiformes present in European waters. The results molecular approaches, pointed that Pleuronectidae and Scophthalmidae were the two most closely related families, and that Soleidae were the most divergent group from the other families studied. Sotelo *et al.* (2001) designed a set of restriction enzymes to provide specific restriction profiles for 24 species of flatfish species commercially available in Europe.

The hydrophobic nature of PAHs and PCBs leads them to be associated with fine sediments (Camacho-Ibar & McEvoy, 1996), thus the benthic lifestyle of flatfishes renders them ideally suited as bioindicator monitoring programs.

1.2.1 Biology of Flounder:

Flounder, *Platichthys flesus*, is widely distributed throughout the coastal waters and estuaries of Europe, and few species are as widespread as the flounder in the European estuarine context, providing an ideal sentinel species for estuarine pollution studies (Eggens *et al.*, 1995; Costa *et al.*, 2002; Stentiford *et al.*, 2003). Flounder is found from the White Sea, south to the Atlantic coast of Portugal, including the Baltic Sea, and relict populations within the Mediterranean Sea (Galleguillos & Ward, 1982; Borsa *et al.*, 1997; Albert *et al.*, 1998; Nissling *et al.*, 2002; Svendäng, 2003).

Flounder live in shallow coastal waters throughout their lifecycle and are rarely found below the 40 m contour line (Albert *et al.*, 1998; Svendäng, 2003). Analysis of their stomach contents reveals that flounder are generalist feeders, targeting indiscriminately any invertebrates in high abundances in the environment (Moore & Moore, 1976; Beaumont & Mann, 1984). Flounders are normally dextrally orientated, as most Pleuronectidae; nevertheless, individuals sinistrally orientated occur in all populations at different percentages: around the British Isles the percentage seems to be around 6%; while in the Baltic, reversed fish can comprise over 35% of the population (Beaumont & Mann, 1984). Orientation is suspected to be related to maternal inheritance (Prof. P. Dando, personal communication).

Natural reproduction is preceded by the adults migrating offshore to fully marine environments from the feeding grounds in coastal lagoons and estuaries. Aggregations occur at particular sites, where the adult stock mass spawns. In the North Sea they seem to overlap

with those of plaice (van der Veer *et al.*, 1991), which are located in the Southern Bight and the Eastern English Channel (van der Veer, 1986). There seems to be a cline in the spawning time throughout the species range (Beaumont & Mann, 1984; Bekhti *et al.*, 1985; Lønning *et al.*, 1988; Rijnsdorp & Witthames, 2005). Highest egg concentrations in the North Sea are in February (van der Land, 1991), but can be detected from December till May (Beaumont & Mann, 1984; van der Veer, 1985).

Under normal open sea conditions, the eggs and larvae drift with residual currents, and approach the coastal zone from where larvae will settle in nursery areas. The pelagic eggs and larvae suffer strong mortality, especially under predation by coelenterates, (van der Veer, 1985), and flounder year-class strength is thought to be controlled by this stage (van der Veer *et al.*, 1991). The larvae migrate into the estuaries and tidal flats near river mouths, and show strong preference for riverine fresh waters (Jager, 1998). Larvae also show a strong rheotaxic behaviour, which further favours the migration into rivers (van der Veer *et al.*, 1991). Young flounder show selective tidal transport (van der Veer *et al.*, 1991), and use the tides to move up the estuaries. Fish are more buoyant in full strength sea water, thus, they *float* (and actively swim) during flood tide and *sink* (and remain in the sediment) during ebb tide (Jager, 1998). The first settlers after metamorphosis appear in brackish-waters with medium-sized sediment or muddy areas by the end of April or middle of May (van der Veer *et al.*, 1991). Flounder can remain in the nursery areas for as long as three years, and then migrate to marine waters in order to breed (van der Veer *et al.*, 1991).

The Baltic Sea is effectively a large brackish water estuary, with surface salinities decreasing from West to East and from South to North, where they can be as low as three practical salinity units (psu). Salinity also varies vertically, with a halocline at 50-70m deep, where salinity is between 11-22psu, however this bottom layer is often hypoxic (Nissling *et al.*, 2002). The Baltic constitutes a particular adaptive challenge: it is inhabited by both marine and freshwater fauna, and both have had to adapt to life and reproduction at low salinity (Andersson *et al.*, 1981; Nielsen *et al.*, 2004; Limborg *et al.*, 2009). Nissling *et al.* (2002) found that salinity was a decisive factor in successful reproduction, as eggs lost their neutral buoyancy at low salinity. It was found that flounder have colonised the Baltic by two different means: 1) flounder in the southern Baltic have modified their sperm activation and egg buoyancy to lower salinities (*ca.* 11psu) and restricted their reproduction to the deep basins with higher salinity; and 2) populations in the northern parts of the Baltic have gone a step further in the adaptation to nearly fresh water conditions (*ca.* 6psu), by laying demersal eggs

in shallow areas. These differences in biological adaptations between northern and southern Baltic flounder were later confirmed to coincide with population structure using neutral genetic markers (Hemmer-Hansen *et al.*, 2007b; Florin & Höglund, 2008). Furthermore, flounder from the North and Baltic Seas also showed genetic divergence at non-neutral genetic markers (Hemmer-Hansen *et al.*, 2007a) and showed adaptive differences in stress response to salinity (Larsen *et al.*, 2007; 2008).

Galleguillos & Ward (1982) found, by the means of allozymes that all populations sampled from the Danish Belt, English Channel, and Irish Sea did not show strong genetic differentiation, having an overall genetic identity of 0.998. As a sequel to Galleguillos & Ward (1982) findings, Borsa *et al.* (1997) using allozymes and mitochondrial DNA, undertook more exhaustive sampling including the Danish belt, Kattegat, North Sea, French Atlantic Coast, and Portugal. Within the Mediterranean basin, the sampling stations extended from the Ebro Delta, the Gulf de Lyon, to the Adriatic and the Aegean. The overall trend was weak geographic differentiation from the Danish Belt Sea to Southern Portugal, while strong divergence was found between Mediterranean and Atlantic populations. The low levels of differentiation on the Atlantic coast were regarded as an indicator of substantial gene flow.

Flounder is now one of the key sentinel species for marine pollution assessment within the European and British context (Eggens *et al.*, 1995; Simpson *et al.*, 2000; Stentiford *et al.*, 2003; George *et al.*, 2004; Sheader *et al.*, 2004; Marchand *et al.*, 2006), but much information on population definition, composition and structure, migration rates between estuaries, and effective population size should be obtained in order to fully exploit the benefits of biomonitoring data.

1.2.2 Biology of Dab:

Dab, *Limanda limanda* are widely distributed, from the Bay of Biscay to the White Sea. The population centre of mass is located in the North Sea, where they are the most abundant flatfish, and third ranking fish after sprat *Sprattus sprattus* and Raitt's sand eel *Ammodytes marinus* (Henderson, 1998). The abundance of dab decreases from the centre of the distribution towards the edges, and dab are four times more common in southern Norway (Karmøy area) than in northern Norway (Lofoten Archipelago) (Albert *et al.*, 1998). Dab are also found in the Irish Sea, Iceland, the Faeroe Islands, and the entrance of the Baltic, where the species is present only as far north as the island of Gotland (Htun-Han, 1978).

The North Sea dab biomass is estimated at around two million tonnes (Daan *et al.*, 1990). In fact, dab numbers are thought to have tripled since the early 70's as a result of reduction of predators, increased food availability due to eutrophication of the North Sea, scavenging of by-catch and trawling disturbance (Kaiser & Ramsay, 1997), and sea temperature rise (von Westernhagen *et al.*, 2002). Although some individuals have been caught at depths of 200m (Bakhsh, 1982), dab generally inhabit waters less than 100 meters deep. For example, in the south eastern North Sea the vast majority of the population are found in shallow sandy substrata less than 30m deep (Henderson, 1998). Dab abundances also seem to vary throughout the year, with higher concentrations in the shallows (20-40m) from July to December (Bakhsh, 1982).

Braber & de Groot (1973) studied the diet of adult and juvenile dab. Several taxa are were found in dab stomach: crustaceans (Mysids, Gammarids, *Crangon, Carcinus* and *Portunus*) and polychaetes (*Lanice, Arenicola, Nereis* and *Pectinaria*). Also prevalent in young dab diet were mollusc siphons and fish eggs. Bryozoans, hydroids and echinoderms were also recorded in minor quantities. There seems to be some seasonality in the feeding patterns of dab, with polychaetes consumed in winter, molluscs in spring, and ascidians in autumn (Ortega-Salas, 1980). However, dab around Anglesey, in the Eastern Irish Sea, exhibit a much higher proportion of echinoderms, especially *Amphiura brachiata*, in stomach contents, to the point of becoming the dominant species in the dab diet (Kaiser & Ramsay, 1997; Seisay, 2001). On the whole, dab are also opportunistic feeders, feeding on the most common prey available in their environment (Hinz *et al.*, 2005).

Like other flatfish, dab take advantage of selective tidal transport and ride the spring and ebb tides to economise the energetic cost of migrations (de Veen, 1978), which they undertake seasonally between coastal feeding grounds and more offshore breeding areas (Rijnsdorp *et al.*, 1992). Overall, the onset of reproductive activity for dab starts in January and ends in September, and is characterised by being particularly extended at most locations (van der Land, 1991). However, as with flounder, dab reproduction is modulated according to location, starting when water temperature reaches 7 or 8°C (Ortega-Salas, 1980). In the southern tip of the geographical distribution, spawning may start as early as January (Henderson, 1998), then peaks in February or March in Southern North Sea (Htun-Han, 1978), in March-April in the German Bight (Rijnsdorp *et al.*, 1992), and progresses upwards towards Norway, where peak spawning is in May-June (Henderson, 1998). Around the British Isles, most eggs are released between February and May. In the Irish Sea, females are ripe from February to May (Ortega-

Salas, 1980; Bakhsh, 1982), with most females spawning in March (Seisay, 2001). Icelandic females reach running condition in May (Jonsson, 1966), and Baltic populations experience spawning peaks in May –July (Nissling *et al.*, 2002). In higher latitudes, breeding time is more tightly constrained, and is later in the year. In the Barents Sea, spawning takes place in mid-July (Henderson, 1998). Individual females spawn for a maximum of six weeks (Htun-Han, 1978), and within locations, older and larger females start the breeding season, while first time spawners follow a little later (Ortega-Salas, 1980). Breeding grounds are usually located offshore, in depths between 20-50 m (Henderson, 1998). The location of several spawning grounds has been estimated by measuring the sea surface egg and larvae concentrations across the North Sea: the German Bight, the north of the Friesian Islands, the southern edge of the Dogger Bank, and off Flamborough Head (Rijnsdorp *et al.*, 1992; Bolle *et al.*, 1994); and the Irish Sea: where higher concentrations are found off Ireland and in the Liverpool and Cardigan Bays (Fox *et al.*, 1997). Some local adaptation to lower salinities has been observed in dab populations from the Baltic; nevertheless, dab distribution does not reach the same coverage as flounder. Successful reproduction occurs occasionally in the Arkona and Bornholm basins only when marine water plumes from the North Sea enter the Baltic (Nissling *et al.*, 2002). Egg size is slightly larger in the Baltic than in the North Sea (Henderson, 1998).

Generally, the eggs undergo passive dispersion and are always the most abundant fish species in pelagic eggs assemblages throughout the species geographical range in February and March (Campos *et al.*, 1994). However, exceptions occur when very cold surface temperatures are prolonged into the spring, when plaice eggs become more common than those of dab (von Westernhagen *et al.*, 2002). Henderson (1998) undertook a comprehensive study on recruitment of dab, and found that the benthic lifestyle starts after metamorphosis at a length of 13-20mm. Larvae do not show rheotaxic behaviour, and are not constrained to estuaries as nursery areas. Settlement occurs in both shallow coastal areas and offshore. Therefore, young dabs are found in a much wider and deeper depth range than young plaice. Temporally, while plaice juveniles are already present in the nursery in April, newly settled dab in the west coast of Scotland do not appear until mid-May, and are uncommon until June (Gibson *et al.*, 2002). In Dutch shallow waters (4-10m), large numbers of 0-group fish are not seen until September/October (Bolle *et al.*, 1994) and stay in high numbers until December (Henderson, 1998). There is a second peak in dab abundance in February, when one year old juveniles enter the shallow waters (Henderson, 1998). Plaice juveniles migrate onshore to shallow waters (*ca.* 0.5m) in order to avoid predation and maximise growth (Gibson *et al.*,

2002), but dab juveniles are rarely caught intertidally. Laboratory experiments with dab juveniles found that maximum growth occurred between 15°C and 18°C, and was halted at temperatures above 22°C (Bolle *et al.*, 1994), so dab may be effectively temperature restricted to deeper waters. A significant negative relationship between winter-spring temperature and young of the year juvenile abundance in autumn has been found throughout the southern half of the species distribution (Henderson, 1998).

Growth rates are also different depending on locality. North Sea fishes have the slowest growth rates, which might be related to food quality and hyper-abundance of dab and competition (Rijnsdorp *et al.*, 1992). Irish Sea dab grow faster than North Sea or Icelandic dab during the first three or four years, after which the trend is reversed (Bakhsh, 1982), limiting Irish Sea fish to a maximum size of 36.7 cm for females and 26.0 cm for males (Seisay, 2001). Southern most populations in the coast of Brittany, experience the fastest growth rates and the largest recorded sizes, with maximum female size of 43.5 cm and maximum male size of 39.0 cm (Deniel, 1990). Dab usually stay in shallow coastal waters until they reach the age of two years when they migrate offshore to join the adult population. In the North Sea first reproduction is at one year for males, and two years for females, at mean lengths of 10 cm and 14 cm respectively (Henderson, 1998). However, Rijnsdorp *et al.* (1992) report later sexual maturation in the North Sea by one year, and in Icelandic populations males are presumed to mature at 2-3 years and females at 3-4 years old (Jonsson, 1966). In the Irish Sea, mean length and age at maturation for males and females were 12 cm and 19 cm total length, and 0.86 and 3.06 years old respectively (Bakhsh, 1982). Brittany females also join the breeding stock at three years of age but are then 27 cm long. Life expectancy also varies throughout the geographical range: in the North Sea, fish as old as 11 years can be found (Henderson, 1998), while in the Irish Sea dab only reach 8 or 9 years (Ortega-Salas, 1980; Seisay, 2001). Dab mortality as by-catch from the plaice fishery is high (Kaiser & Spencer, 1995), and juvenile mortality rates seem to be comparable in both East and West British coasts (Henderson, 1998).

It is generally recognised that dab can be subdivided into several local stocks depending on their spawning time and location, size, growth rate, life expectancy, and morphometric variation such as relative size of head and number of rays in dorsal and anal fins (Bakhsh, 1982; Deniel, 1990; Rijnsdorp *et al.*, 1992; Henderson, 1998). Yet their pelagic larval stage may result in strong gene flow between these populations and the possibility of panmixia arises. Whether these morphometric and life-history-traits differences emerge due to genetic

differentiation or are a product of phenotypic plasticity modulated by the environment is not yet known. A resolution to this issue is needed if dab are to be successfully managed or used in biomonitoring programs (Lyons *et al.*, 2000).

Apart from the current thesis and studies on pollution effects on DNA integrity (Everaarts, 1995; Lyons *et al.*, 2000; CEFAS, 2003b; Lyons *et al.*, 2006), relatively few published studies have used genetic tools in dab. The phylogenetic relationship of dab with other European flatfish species has been resolved (Verneau *et al.*, 1994; Exadactylos & Thorpe, 2001; Sotelo *et al.*, 2001; Espiñeira *et al.*, 2008), and the number of chromosomes (2n=26) has been studied (Di & Knowles, 1994). Furthermore, a line of research has followed the isolation of genes in dab involved in either oncogenic processes, *Ras* genes (Rotchell *et al.*, 1995), or tumour suppression, retinoblastoma genes (Du Corbier *et al.*, 2005; Rotchell *et al.*, 2009).

1.3 Biomarkers:

Measureable entities such as morphological and biochemical responses and compensatory mechanisms exhibited by an organism after exposure to contaminants are used as biomarkers. A good biomarker is one which is able to elucidate whether an organism has been exposed to pollutants, if the pollutants have been incorporated into the body tissues, and if a toxicological response has been triggered (Everaarts, 1995; Wu *et al.*, 2005). Current standard biomarkers include: 1) assessment of DNA adducts; 2) histopathology of liver tissues; 3) quantification of CYP1A or EROD activity; and 4) chemical analysis of low molecular weight PAHs excreted into bile as metabolites.

1.3.1 DNA Adducts:

DNA is susceptible to damage to any of its parts: the phosphodiester backbone, the ribose sugars, or the purine or pyrimidine bases (reviewed by Carajaville *et al.*, 2003). The damage can be either at the sequence of nucleotides level or at the structural level. The former one will have mutagenic effects; while the latter will disable replication and transcription.

If damage has been caused at the DNA sequence level, then mutations that affect genes controlling normal cell division, differentiation and apoptosis, can lead to neoplasms (Rotchell *et al.*, 1995; Baumann, 1998; Du Corbier *et al.*, 2005; Rotchell *et al.*, 2009). A neoplasm is an abnormal tissue growth, which can go through a series of stages: initiation, promotion, and progression. At this final stage, further mutations can trigger the transformation of the benign neoplasms into a malignant invasive cancer (Carajaville *et al.*, 2003).

Damage can instead be initially structural: electrophilic chemicals can form covalent bonds with DNA. Some inert compounds, among them PAHs, are metabolised by the organism into an electrophilic state, thus rendering them prone to form the same bonds. These covalent bonds are known as DNA adducts, and may entail inhibition of DNA transcription or blockage of DNA replication, therefore behaving mutagenically, and ultimately resulting in tumorigenic (Hawkins *et al.*, 1988; Baumann, 1998; Reichert *et al.*, 1998; Carajaville *et al.*, 2003). Pollution exposure has been successfully detected with the use of DNA adducts or DNA strand breakage (Everaarts, 1995; Theodorakis & Shugart, 1997; 1998; Theodorakis *et al.*, 1997; 1999; Neuparth, 2004) and the technique is commonly employed in biomonitoring surveys of flatfish (Lyons *et al.*, 2000; CEFAS, 2003b; Lyons *et al.*, 2006).

1.3.2 Liver Histopathology:

As discussed under the DNA adduct section, anthropogenic pollutants can lead to the formation of tumours in several ways. Therefore, fish liver can be analysed by histology to find hepatic foci of cellular alteration (FCA), benign adenomas and malignant hepatocarcinomas (Feist *et al.*, 2004), which can all be used as a biomarkers of lifelong pollution exposure. These types of data can then be used in comparative studies between locations or in time-series (Simpson *et al.*, 2000; Stentiford *et al.*, 2003; Feist & Stentiford, 2005; Stentiford *et al.*, 2009).

1.3.3 EROD Activity:

Organisms have developed mechanisms to detoxify themselves, of which the Mixed Function Oxygenase (MFO) enzyme system is the primary detoxification pathway for contaminants such as PCBs and PAHs. These mechanisms are induced by exposure to pollutants. The terminal component of the MFO system is known as Cytochrome P4501A1 (CYP1A1), which catalyses hepatic ethoxyresorufin-O-deethylase (EROD) (Eggens *et al.*, 1995). Both CYP1A1 and EROD levels in liver cells has been used extensively as an indicator of exposure to PCBs and PAHs (Wirgin *et al.*, 1991; Förlin & Celander, 1993; Larno *et al.*, 2001; CEFAS, 2003c). However, EROD levels do not have a simple relationship with DNA strand damage (Lyons *et al.*, 2000; Larno *et al.*, 2001), so information collected with one marker is not necessarily comparable to data from another marker.

1.3.4 Bile metabolites:

Fish readily absorb PAHs from the environment, which are processed in the liver and then may be excreted through the bile. Empirical studies have found that measurement of bile metabolites is equally sensitive to and complements the analysis of EROD activity (Gagnon &

Holdway, 2000). Thus, fish bile is considered to be representative of previous exposure to petroleum derivatives and used in biomonitoring programmes evaluating flatfish health (Neall *et al.*, 2003).

1.3.5 New biomarkers of exposure to pollutants:

Apart from the biomarkers mentioned above, there is an interest to detect pollution adaptation, thus, efforts are being diverted towards developing a suite of markers sensitive to gene variability (George *et al.*, 2004).

Detoxification-related genes, as all genes, are subject to mutations and may show different levels of variability between species (Reichert *et al.*, 1998), between populations, and ultimately between individuals (George & Leaver, 2002). One of the genes within the UGT family, UGT1B1, isolated from plaice (*Pleuronectes platessa*), shows high levels of polymorphism between individual plaice after restriction digestion (George & Leaver, 2002). These expressed genes are vital to the survival of the individual, and mutations at this level may have relevant effects on the demography and age class distribution of natural populations. Other detoxification mechanisms that might show variability include: the UDP-glucuronosyltransferase (UGT), which neutralize lipophilic toxic wastes and pollutants (George & Taylor, 2002); glutathione-S-transferase (GST), which ligate many toxic compounds to the cysteine sulphur of glutathione, and neutralize lipid peroxidation products (Martinez-Lara *et al.*, 2002); epoxide hydrolase (EH), which is suspected to help in the excretion of PAHs (Willet *et al.*, 2000); REL family proteins bind with kB factors and enable genes encoding for cytokines, and acute phase proteins among others (Schlezinger *et al.*, 2000); Orosomucoid 1 and α-1-acid glycoprotein (AGP) were expressed in intoxicated rat liver (Kondraganti *et al.*, 2005). If sufficient polymorphism can be found in a number of these detoxification genes, they could be used as adaptive markers for population analysis.

Much interest has recently been focussed towards identifying differentially expressed genes related to pollutant exposure. George *et al.* (2004) quantified expression of CYP1A mRNA by the use of real time quantitative-PCR, and detected slight differences between polluted and reference sites. Sheader *et al.* (2004; 2006) applied suppressive, subtractive hybridisation (SSH) to flounder from exposed and non-exposed environments, and isolated 284 mRNA sequences with potential differential expression. Of these sequences, 84 could be cautiously identified to other published sequences, of which some were thought to have detoxifying qualities. Marchand *et al.* (2006) employed the same technique, and compiled 256 expressed genes: fourteen were tested by reverse-transcriptase PCR on flounder from contaminated

and reference sites and eight genes were differentially expressed when exposed to high or low concentrations of an herbicide cocktail. Assessment of the technique was then taken to the field: flounder from polluted environments showed up-regulation of seven genes during the summer months matching high herbicide concentrations, compared to winter months when herbicide levels are low. Flounder from low pollution estuaries showed little change in expression of these genes throughout the year. Three of these genes were energy-related: NADH dehydrogenase, Cytochrome *c* oxidase, and ATP synthetase, suggesting that flounder in polluted environments require more energy. Three other genes have detoxifying qualities: lipocalin-type prostaglandin D synthetase (L-PGDS), betaine homocysteine methyltransferase (BHMT), and carboxylesterase, which are involved in the proteolytic degradation of herbicides. The other genes, such as Hepatocyte growth factor and Elongin C, could indicate potential carcinogenesis. SHH proves as a powerful technique to establish new differentially expressed detoxification-related genes, thereby providing a new generation of biomarkers for assessing contamination of aquatic environments.

1.4 Evolutionary Ecotoxicology:

There is an increasing interest in tackling the effects of pollutants in populations, by merging ecotoxicology and population genetics (D'Surney *et al.*, 2000; Bickham *et al.*, 2000; Belfiore & Anderson, 2001; Theodorakis, 2001). The effects of pollution go beyond the individual, and could have extensive consequences at the population and species level. The pressures imposed by environmental pollution on individual organisms disappear when pollution ceases, however, the effects on the gene pool are of a more permanent nature.

In the case of bioindicator marine fish species, there are four plausible outcomes of genetic variability (and consequences for biomonitoring programmes) after selection by pollution. These vary depending on the intensity of the selection pressure and the patterns of connectivity between exposed and non-exposed populations (modified from Baker *et al.*, 2001):

1) Evolutionary scenario A:

Populations are mainly self-recruiting with little gene flow from other populations. Isolation provides the potential for local adaptation. Genes providing pollution tolerance become widespread in polluted areas. Consequences: Potential loss of genetic variability and directional change towards pollution-tolerant alleles in detoxification genes. Increased pollution tolerance may bias biomarker data.

Figure 1.1: Evolutionary scenario A:

LEGEND: Three time points are represented on top of each other, beginning at the top. The two sides of the box represent two populations, while the lines in between them represent barriers to gene flow either by larval recruitment (grey arrows) or adult migration (black arrows). The sizes of the cartoon fish represent age. The different colours of the cartoon fish represent genetic diversity. One gene (red) grants tolerance to pollution (which may in homozygous or heterozygous state). Pollution increases in the right hand population over time from top to bottom as brown background. The black crosses indicate pollution-induced mortality.

2) Evolutionary scenario B:

Adult fish are sedentary and under strong local pollution selection, but larval dispersal is high. New recruits may be genetically variable for some time, but adult fish are the survivors of strong pollution-related mortality. Consequences: Biomarker data needs to be interpreted accordingly, as older pollution-tolerant fish may show less detrimental effects compared to the young ones.

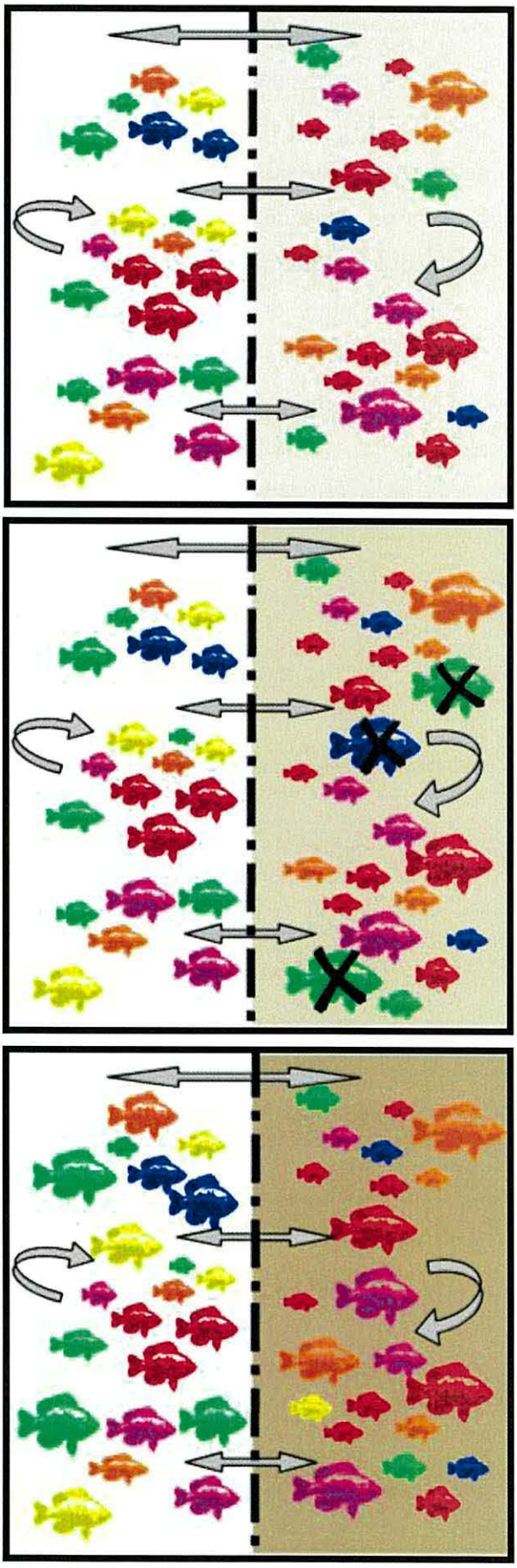


Figure 1.2: Evolutionary scenario B:

LEGEND: As in evolutionary scenario A.

3) Evolutionary scenario C:

Strong adult migration and larval dispersal result in no population structure (populations cannot be distinguished genetically). Consequences: No detectable pattern of selection at the population level. Local adaptation is hindered by the continuous influx of foreign genotypes. Under such circumstances, biomarker data may not be representative of capture location.

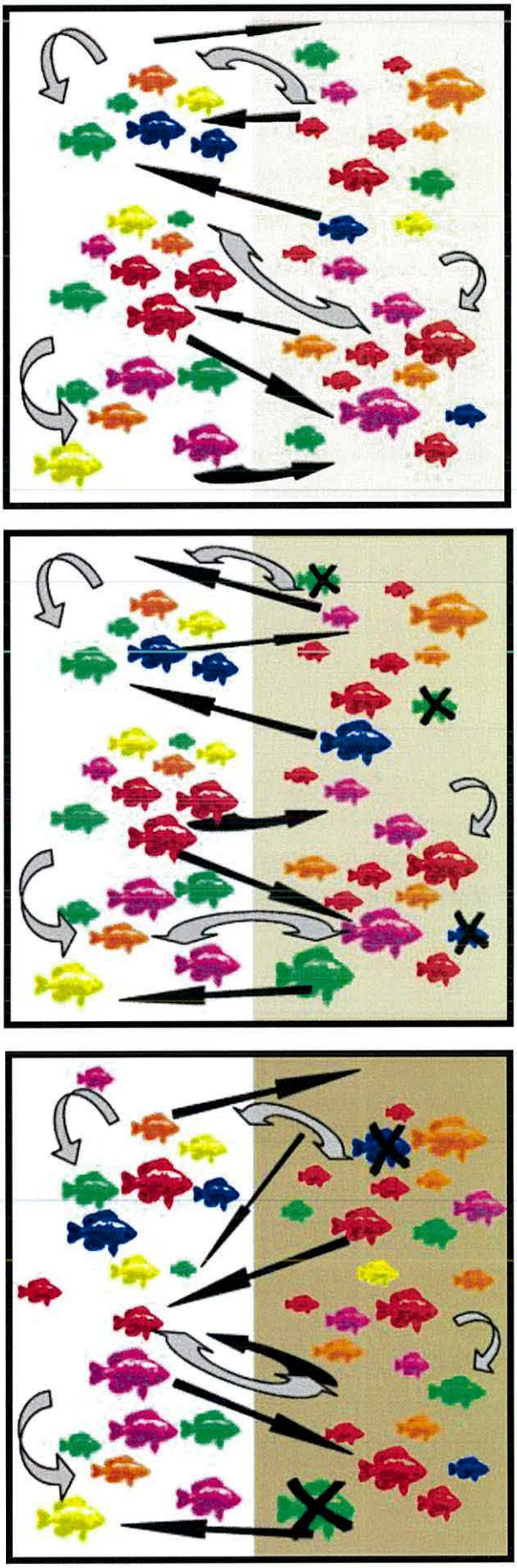


Figure 1.3: Evolutionary scenario C:

LEGEND: As in evolutionary scenario A.

4) Evolutionary scenario D:

Pollution is so high in some areas that it prevents reproduction. Polluted areas become sink populations receiving larvae and/or adults from other less polluted areas (sources). Consequences: Sink populations will show genetic signals of immigration and lack of temporal stability, which are both indicative of the heavy pollution impact on the environment and taxa.

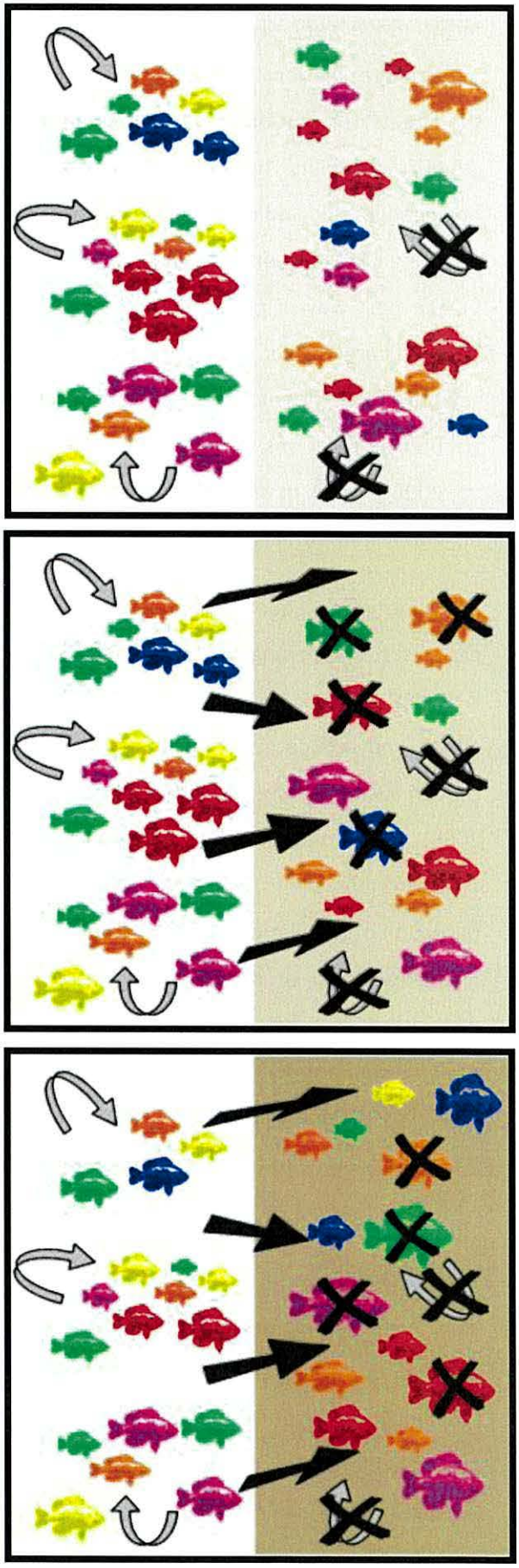


Figure 1.4: Evolutionary scenario D:

LEGEND: As in evolutionary scenario A.

These four hypotheses are not mutually exclusive and combinations are likely. Furthermore, it might prove very difficult to distinguish between such hypothesis and natural variation between populations (Theodorakis & Shugart, 1997; Baker *et al.*, 2001; Roark *et al.*, 2001). If the effects of pollution on a population are to be assessed, the other natural factors affecting genetic variability (mutation, migration, genetic drift and selection) and the species phylogeographic and demographic history need to be well understood beforehand (Cronin & Bickham, 1998; Staton *et al.*, 2001). Long term pre- and post-exposure genetic analysis would be the best way of detecting the pollution effects on the gene pool, though such natural scenarios are scarce. The potential selective effects of pollution have been demonstrated in several empirical studies based on the dynamics of genetic diversity in polluted and non-polluted populations, both in controlled replicate experimental populations (Gardeström *et al.*, 2006; Nowak *et al.*, 2009) and in wild populations (Theodorakis & Shugart, 1997; Peles *et al.*, 2003; Theodorakis *et al.*, 2006; Bourret *et al.*, 2008).

The genetic consequences of pollution cannot be taken lightly. Low levels of genetic diversity can lead to inbreeding, which can have very different and unpredictable consequences (Amos & Balmford, 2001; England *et al.*, 2003). On one hand, inbreeding can purge deleterious alleles and reduce genetic load, enhancing the chances of survival in a particular habitat, though such processes do not necessarily result in enhanced overall fitness (Rowe & Beebee, 2005). On the other hand, inbreeding can raise levels of homozygosity, increase the frequency of deleterious alleles, and ultimately elicit a *mutational meltdown*, where a vicious loop of reduction of population size and inbreeding, reciprocally encouraged by each other, incur in a final extirpation or extinction (Amos & Balmford, 2001; Beebee & Rowe, 2004). Inbred populations of guppy *Poecilia reticulata*, invariably show reduced salinity tolerance (Shikano *et al.*, 2001), and highly heterozygous fishes experience higher levels of detoxification rates and less DNA damage after exposure to chemical pollutants (Larno *et al.*, 2001; Maes *et al.*, 2005), or are able to tolerate higher levels of pollutants (Bourret *et al.*, 2008).

Theodorakis *et al.* (1997) found a positive correlation between levels of radiation, DNA structural damage (in the form of strand breaks), and reduced fertility and higher numbers of abnormal embryos in mosquitofish, *Gambusia affinis*. If radiation is causing DNA structural damage and the latter is related to reduced fertility, a population exposed to high levels of radioactive pollution will experience reduced reproductive output, and have profound effects on the population's viability (Theodorakis, 2001). Individuals from radiated and control populations varied in randomly amplified polymorphic DNA (RAPDs) profiles, and certain

bands were particularly common in individuals from radioactive sites (Theodorakis & Shugart, 1997). Translocation of fish from control sites to radioactive ones resulted in differential survival of fish: those with the same RAPD bands as the radiated population survived (Theodorakis *et al.*, 1999). Organochlorine residues may be an important source for malformation of dab embryos, and Dethlefsen *et al.* (1996) found a significant reduction in the frequency of abnormalities as the concentration of organochlorine in the North Sea was lowered since 1987.

Long-term environmental contamination pressures will have profound effects on the gene pool of natural populations, and in light of the negative synergistic effects between pollution and climate change (Schiedek *et al.*, 2007), it becomes imperative that the evolutionary processes involved in the long-term exposure to anthropogenic contaminants are understood. Polymorphic molecular markers in the context of population genetics emerge as leading candidates to tackle such questions related to predictive response.

1.5 Introduction to Population Genetics:

1.5.1 Population genetics in the marine environment:

One of the key roles of geneticists in fisheries and environmental studies is the identification of self-contained groups of individuals which comprise a definable genetic pool or breeding unit, distinguishable from other similar groups of individuals of the same species (Carvalho & Hauser, 1994; Carvalho & Hauser, 1998; Thorpe *et al.*, 2000; Ward, 2000; Hauser & Carvalho, 2008). Genetic information is needed to define the population structure of a species, to establish breeding and conservation units (Fraser & Bernatchez, 2001; Schwartz *et al.*, 2007), and to understand the genetic variability and the distribution of that genetic variability present within a species (Waples & Gaggiotti, 2006).

In the fisheries field the identification of reproductively isolated units, known as *stocks*, is of particular importance (Carvalho & Hauser, 1994; Ward, 2000). If a fishery is to be successfully managed, attention must be paid at the level of sub-structuring of the species to be managed. Furthermore, populations may be adapted to local conditions (Carvalho, 1993; Naish & Hard, 2008), and if a locally adapted stock is overexploited, it might not be possible to replenish the stock with individuals migrating from adjacent areas.

Marine fish usually support large census populations, have pelagic larvae and are capable of extensive migration, which combined create the prediction of panmixia, whereby they are expected to show genetic homogeneity over large sections of the sea (Ward *et al.*, 1994;

Waples, 1998). Nevertheless, there is increasing evidence that marine fishes experience higher levels of genetic structuring than previously expected (Andersson *et al.*, 1981; Pogson *et al.*, 1995; Magoulas *et al.*, 1996; Lundy *et al.*, 1999; Ruzzante *et al.*, 1999; Hutchinson *et al.*, 2001; Mattiangeli *et al.*, 2002; O'Reilly *et al.*, 2004; Mariani *et al.*, 2005), and that the level of gene flow needed to overcome the urge for local adaptation may have been underestimated (Hemmer-Hansen *et al.*, 2007a; Larsen *et al.*, 2007).

1.5.2 Population genetics basic concepts:

Thomas Hardy and Wilhelm Weinberg established the basis for population studies based on gene frequency at the beginning of the last century (Hedrick, 1999a; Beebee & Rowe, 2004). The Hardy-Weinberg equilibrium (HWE) model states, that, if two alleles of the same locus, A and B, are present in frequencies p and q (where $p + q = 1$), then the frequencies of homozygotes for A and B are p^2 and q^2, while heterozygotes (AB) will be present at $2pq$ (furthermore, $p^2 + q^2 + 2pq = 1$). If a group of individuals is randomly mating and is not subject to mutation, selection, genetic drift, nor migration, then gene frequencies are expected to remain unchanged from one generation to the next. Natural populations hardly ever meet the ideal requirements for the HWE model, and measurement of the deviations from the model can be used to investigate population substructuring. Wright's *F*-statistics (Wright, 1943; 1951) give estimates of deviations from expected heterozygote frequency (*F*), and are calculated as:

$$F = \frac{H_E - H_O}{H_E}$$

Where, H_E and H_O are the expected and observed heterozygosity. Departures from the HWE model expectations are valuable as they may reveal substructuring of the population. To quantify the departure, three different hierarchical levels are used. F_{is} relates individuals to subpopulation, F_{it} relates individuals with total population, and F_{st} relates subpopulations with total population. The latter is the most widely used, and it evaluates differences at the level of subpopulations from the total population. It is calculated as follows:

$$F_{st} = \frac{H_t - H_s}{H_t}$$

Where, H_t is the total population heterozygosity, and H_s is the within-population heterozygosity (Weir & Cockerham, 1984). In undifferentiated equilibrium populations *F*-statistics are all zero (provided the HWE assumptions are met). The greater the departure

from zero the stronger the sub-structuring of the populations. Another way of evaluating genetic relationships is using Nei's Identity (*I*) (Nei, 1972). Based on allele frequency in each population, it ranges from one (identical allele frequencies) to zero (no alleles shared). If p_1 and p_2 are allele frequencies in subpopulation 1 and 2, *I* is calculated as:

$$I = \frac{\Sigma(p_1 p_2)}{\sqrt{(\Sigma p_1^2)(\Sigma p_2^2)}}$$

From Nei's identity, Nei's genetic standard distance (D_S) can be calculated as:

$$D_S = -Ln(I)$$

There are four key factors affecting the distribution of genetic diversity in populations:

1) Mutation: is the source of genetic diversity and happens randomly across the DNA sequence creating variants (alleles) at certain locations of the DNA (Locus/Loci).

2) Selection: some of these mutations may entail negative effects on the organism (i.e. loss of function), and thus will be selected against and disappear. The result is an increase or fixation of the alternative allele, until new alleles are created.

3) Genetic drift: once there are a number of alleles at a locus, fortuitous changes in reproductive output between individuals will lead to changes in allele frequencies between generations. These changes can range from minimal to strong depending on the relative number of breeding individuals. Differences in genetic drift between isolated populations will generate differences in allele frequencies between those populations.

4) Migration: entails the translocation and successful interbreeding of individuals between populations. Therefore, alleles are exchanged between populations. If migration is high enough to overcome genetic drift and local selective forces, the allele frequencies will be homogenous between populations. The latter case is known as panmixia.

These opposing forces are the drivers behind the creation of population structuring, as the rate of mutation, selection, genetic drift and migration may be different in varying locations of the species range. The process is known as divergence, and allows formation of

subpopulations. When gene flow between two populations of the same species is prevented by some geographical impediment, such as a mountain range or large extensions of water, the populations experience vicariance, and may ultimately divert into different species.

1.5.3 Molecular Markers:

The basis of molecular ecology is to assess the variability of genomic information contained within individual organisms that makes them unique when compared to other organisms. The information is conserved as DNA, which is inherited by the offspring, and may be translated into proteins. In order to assess the genetic diversity found among individuals or populations of a particular species, molecular markers can be identified and studied. Molecular markers are protein or small nucleic acid molecules which have variations in their DNA sequence (i.e. they are polymorphic) and, thus, have at least two alleles or variants. Diploid eukaryotic individuals may then be scored as homozygous (two identical alleles), or heterozygous (two different alleles). If molecular markers are neutral, they are considered to be representative of the processes affecting the whole genome, such as isolation or migration between subpopulations. Conversely, if they behave independently from the rest of the genome they may be under the influence of other external factors such as positive or balancing selection (Schlötterer, 2002). A brief overview of two genetic markers relevant to the current thesis is now provided:

1.5.3.1 Allozymes:

The first molecular markers to be widely used were allozymes, which are polymorphic enzymes varying in their amino acid sequence, and can be scored according to their mobility in an electric field. Allozymes show Mendelian inheritance and considerable protein polymorphism has been revealed between organisms from different species (Andersson *et al.*, 1981; Blanquer *et al.*, 1992; Verneau *et al.*, 1994; Exadactylos & Thorpe, 2001) or among populations within species (Galleguillos & Ward, 1982; Blanquer *et al.*, 1992; Verspoor, 1994; Kotoulas *et al.*, 1995; Bembo *et al.*, 1996; Borsa *et al.*, 1997; Bouza *et al.*, 1997; Exadactylos *et al.*, 1998; Foss *et al.*, 1998; Giæver & Stien, 1998). However, redundancy of the genetic code prevents detection of substantial amounts of genetic variation, such as silent mutations or synonymous nucleotide substitutions (Beebee & Rowe, 2004). Consequently, classical markers have sometimes overlooked genetic divergence between populations, mainly due to low levels of polymorphism and heterogeneity found with allozymes.

1.5.3.2 Microsatellites:

Microsatellites, also known as simple sequence repeats (SSRs), are tandemly repeated short sequences, mainly dinucleotides, typically present in non-coding parts of the genome (Jarne & Lagoda, 1996; Chistiakov *et al.*, 2006; Li *et al.*, 2002; Oliveira *et al.*, 2006). Their high polymorphism in the form of variation of the number of repeats, putative neutrality as they are not translated into proteins, and codominance, where heterozygotes show both alleles as two bands, have rendered them very useful in ecological management (Schwartz *et al.*, 2007; Waples *et al.*, 2008), forensic analysis (Primmer *et al.*, 2000; Renshaw *et al.*, 2006), and food production issues (Liu & Cordes, 2004; Chistiakov *et al.*, 2006).

Microsatellite variability may arise in several ways, but the most basic source of polymorphism is by replication slippage when the polymerase encounters a long set of repeated short sequences (Viguera *et al.*, 2001; Li *et al.*, 2004). Basically the polymerase gets confused with how many copies are tandemly arranged, and adds or removes a copy. As microsatellites are normally located in non-coding regions, the mutation may be accepted, and a new allele at that locus is created. Microsatellites have a natural tendency towards elongation (Ellegren, 2002b; Seyfert *et al.*, 2008), and longer microsatellite alleles are more prone to mutation than shorter ones (Seyfert *et al.*, 2008; Brandström & Ellegren, 2009), though, as the microsatellite increases in size it becomes unstable, and mutations breaking the repeated sequence may appear, splitting the microsatellite into shorter sections (Ellegren, 2002a) and reducing its mutation rate (Brandström & Ellegren, 2009). Being composed of repeated elements, microsatellites are expected to follow a stepwise mutation model (SMM) in which new alleles emerge by the incorporation or deletion of a single extra repeated element (Kimura & Ohta, 1978). The mutation model followed by microsatellites has attracted considerable attention, as it affects the way estimators of genetic differentiation are calculated (Goldstein *et al.*, 1995a; Shriver *et al.*, 1995; Slatkin, 1995; Michalakis & Excoffier, 1996). However, the general conclusion is that, in most cases, differentiation can be safely estimated assuming an Infinite Allele Model (Balloux & Lugon-Moulin, 2002; Estoup *et al.*, 2002a), and recent studies suggest that generation of new alleles in microsatellites involving more than one repeat element are more common than previously thought (Seyfert *et al.*, 2008).

Microsatellites are amplified from isolated nuclear DNA through PCR, to which oligonucleotides complementary to the flanking regions of the SSR are added. The DNA and PCR mixture goes through several cycles (generally 30) of denaturation (92°C), annealing (temperature depending on the length of the oligonucleotide primers, but generally between

50°C and 60°C), and elongation (72°C), where the SSR is amplified exponentially. If the oligonucleotide primers are marked, either radioactively or by fluorescence, then the size of the SSR can be scored on electrophoresis gels or capillaries (Beebee & Rowe, 2004).

Microsatellites have revealed and deepened the knowledge of population polymorphism and structure where other markers, such as allozymes, have not been able to or showed reduced structure (Ruzzante *et al.*, 1999; Shaw *et al.*, 1999; Ball *et al.*, 2000; Launey *et al.*, 2002; Bernal-Ramírez *et al.*, 2003; Knutsen *et al.*, 2003; Nielsen *et al.*, 2004; Hemmer-Hansen *et al.*, 2007b; Suk & Neff, 2009a). Nevertheless, some studies have also reported better structure resolution with allozymes than with microsatellites (Lemaire *et al.*, 2000; De Innocentiis *et al.*, 2001; Olsen *et al.*, 2004; Gosling *et al.*, 2008). Microsatellites have recently received some attention from the ecotoxicology research community (Brown *et al.*, 2001; Dimsoski & Toth, 2001), and are expected to provide important information on the effects of pollutants on population structure and genetic variability.

There are many microsatellites developed for European flatfish species (Coughlan *et al.*, 1996; McGowan & Reith, 1999; Iyengar *et al.*, 2000b; Iyengar *et al.*, 2000a; Watts *et al.*, 2001; Hoarau *et al.*, 2002a; Funes *et al.*, 2004; Casas *et al.*, 2005; Garoia *et al.*, 2005; Pardo *et al.*, 2005). However, despite the fact that microsatellites developed from one species can sometimes be amplified in related species (Primmer et al. 1996), using cross-species amplified loci can be problematic (Oliveira *et al.*, 2006) and best results are obtained when markers are developed directly from the target species, especially when large numbers of markers are sought.

1.6 Summary and Aims of the thesis:

Impacts of anthropogenic pollutants on natural environments can be assessed by measuring biomarkers such as tumorigenic processes in the liver, DNA adducts, or EROD activity on flatfishes, namely flounder and dab. However, our understanding of the population structure, connectivity between populations, and temporal stability of such structure in flounder and dab is limited. These aspects of the bioindicator species ecology need to be understood for a correct interpretation of the biomonitoring data.

There were two aims for the thesis: 1) to understand the patterns of connectivity between biomonitoring sampling locations and evaluate the possible effects of individual movement; and 2) two evaluate the potential for local adaptation to pollution, and if so whether evidence of selection could be found.

Therefore the objectives of the thesis were:

1. To develop a suite of microsatellite markers for dab and flounder: therefore the isolation and characterisation of the first 30 microsatellite markers for dab and a new set of 28 novel microsatellites for flounder are described in Chapters 2 and 3.

2. To develop a low-cost high-throughput technique: In order to maximise the number of loci and individuals genotyped a novel labelling technique aimed at reducing the costs of testing and genotyping large number of microsatellite markers is described in Chapter 4. Three multiplex reactions for dab are also described here.

3. To evaluate signals of pollution-induced selection: Gene-linked microsatellite markers derived from a pollution-induced expressed sequence tags library are developed and tested for signals of selection in dab afflicted with liver diseases in Chapter 5.

4. To evaluate the best estimator of differentiation: Several new and traditional estimators of genetic differentiation are compared empirically in Chapter 6.

5. To examine the population structure of dab: Several biomonitoring stations of dab are sampled and individuals genotyped. The genetic data are then analysed with several software packages and the results reported in Chapter 7.

6. To evaluate the significance of the genetic data in the biomonitoring context: The results of the genetic analysis are integrated with the biomarker of pollution exposure data. Findings are reported in Chapter 8.

Chapter 2: Development of 30 microsatellite markers for dab (*Limanda limanda* L.): a key UK marine biomonitoring species

2.1 Abstract:

Dab (*Limanda limanda*) are the principal target fish species in offshore biomonitoring programmes in the UK; however detailed knowledge of genetic structure and connectivity among sampling locations is unavailable. Here, the isolation and characterisation of 30 polymorphic microsatellite loci for dab is described. The number of alleles per locus ranged from 2 to 42, with observed heterozygosities ranging from 0.089 to 1. These loci will enable high resolution of genetic population structure and dynamics of dab around the British Isles.

2.2 Introduction:

Flatfishes are ideal indicator species for assessing the biological effects of contaminants in the marine environment, and in the UK the dab (*Limanda limanda*) is studied in annual monitoring programmes (CEFAS, 2005). Although an extensive data base exists on the assessment of individual consequences of pollution exposure (Lyons *et al.*, 2000), information on the genetic structure of dab and population connectivity is limited, both of which are important to correctly interpret biomonitoring data. Furthermore, the potential evolutionary processes in populations displaying elevated levels of disease, or exposed to high levels of pollutants, have not been addressed. Here, the isolation and characterisation of 30 novel polymorphic microsatellite loci for dab, which can be used to analyse the genetic structure of dab populations, is described.

2.3 Methods, Results and Discussion:

A microsatellite enriched genomic library was constructed following a subtractive hybridization protocol (T.C. Glenn, personal communication; www.uga.edu/srel/DNA_lab/Msat_Easy_Isolation_2000.rft). Dab for library development were collected from the Irish Sea and fin clips stored in 100% ethanol until processed. Genomic DNA was extracted using a phenol-chloroform protocol, and 2 µg of DNA was then simultaneously digested with *Rsa I* restriction enzyme (NEB) and ligated to double stranded SNX linkers (SNX-f:

5'CTAAGGCCTTGCTAGCAGAAGC and SNX-r: 5'pGCTTCTGCTAGCAAGGCCTTAGAAAA). Success of the ligation reaction was checked by PCR using single stranded SNX linkers as primers.

Four biotin-labelled microsatellite motif probes $(AG)_{12}$, $(AC)_{13}$, $(ACAG)_6$ and $(AGAT)_8$ were hybridized to the PCR products. Streptavidin Dynabeads (Invitrogen) were used to capture the microsatellite-containing DNA fragments which were eluted in TLE (10mM Tris, 0.1mM EDTA, pH 8.0), amplified by PCR, and PEG (Polyethylene glycol) precipitated. Subsequently, amplicons were A-tailed and ligated into pCR TOPO® Vectors (Invitrogen) and transformed into One Shot TOP10® competent cells (Invitrogen). Recombinant colonies were identified on Luria-Bertani agar plates by ampicillin resistance and disruption of the β-galactosidase gene. Microsatellite presence was evaluated on 960 colonies by PCR amplification with M13 F primer and a mixture of non-biotinylated microsatellite probes. 251 positive amplicons were sequenced by Macrogen Inc. (Korea), and the ensuing sequences edited, analyzed and checked for duplicates with BIOEDIT (Hall, 1999). Enrichment efficiency was high at 87.5%. Primers were designed on either side of 58 putative microsatellites using PRIMER3 (Rozen & Skaletsky, 2000), and tested for successful amplification at several annealing temperatures on 3% TBE agarose gels. Forward primers of pairs reliably amplifying on several individuals were then ordered M13-tailed at the 5' end (Schuelke, 2000). Nested PCRs with forward-tailed primer, reverse primer, and FAM™-labelled M13-tail oligos were used for genotyping. PCR cocktails of 10 µl final volume contained around 20 ng of DNA, 1x GoTaq® Flexi buffer (Promega), 1.5 mM $MgCl_2$, 125 µM dNTP, 0.1 µM Forward-tailed primer, 0.5 µM of Reverse primer, 0.5 µM of FAM™-labelled M13-tails, and 0.5 U GoTaq® DNA polymerase (Promega). PCRs were carried on a BioRad Tetrad2® Peltier Thermal Cyclers and the thermocycling programmes were as follows: an initial denaturation phase of 3 min at 95°C, followed by 13 cycles of 30s at 95°C, 45s at the forward primer annealing temperature (Table 2.1), 60s at 72°C, then 31 cycles of 30s at 94°C, 45s at 50°C, 60s at 72°C, and finishing with a 30min extension phase at 72°C.

Two samples of 24 dab from two locations, North Sea (55°17'59.96''N; 2°53'45.81''E) and Irish Sea (54°30'42.48''N; 3°47'37.68''E), were genotyped on an ABI 3130xl Genetic Analyzer (Applied Biosystems) with an internal size standard (GeneScan™ LIZ-600®). Allele sizes were scored with GeneMapper® Software 4.0. Thirty primers produced polymorphic bands at the expected sizes (Table 2.1). Genotypes were analyzed with GENETIX (Belkhir *et al.*, 1996-2004) and GENEPOP V4.0 (Rousset, 2008), where polymorphism varied from 2 to 42 alleles with an average of 15 alleles per locus. Observed heterozygosity ranged from 0.083 to 1. Significant

deviations from the Hardy-Weinberg Expectations in the form of heterozygote deficiencies were found in DAC4-34, DAC5-70, DAG2-15, DAG2-22, and DAG4-91 in either or both populations, suggesting the presence of null alleles in these markers. Marker DAC5-21 was highly similar/homologous to the *Hippoglossus hippoglossus* microsatellite Hhi61IMB (GenBank: EF569094), and parts of the sequence of DAC2-15 were highly similar to the Potassium Chloride transporter gene (BC136157.1; blastn value=$2e^{-11}$). Although significant linkage disequilibrium was found in one of two populations between several loci, no pair of loci was significantly linked for both populations, suggesting that linkage is likely to be an artefact of small sample sizes. Cross-species amplification on eleven European flatfish species was tested using the same parameters as for dab (Table 2.2). These markers will prove invaluable for the description of genetic population structure, connectivity and demographics of dab around the British Isles. Furthermore, they will enable more accurate interpretation of biomonitoring data, and provide a neutral genetic background with which adaptive genetic markers can be compared.

Table 2.1: Characteristics of 30 microsatellite loci for dab, *Limanda limanda*.

Characterization of 30 microsatellite loci isolated from *Limanda limanda* in two populations. Motif = repeat sequence of the isolated clone; Ta = Annealing temperature; N= number of individuals successfully amplified (out of 48). Na = Number of alleles; Range = allele size range; Ho = Observed Heterozygosity; He = Expected Heterozygosity; p = associated probability value of conformation with Hardy-Weinberg Equilibrium (HWE). Bold p values indicate significant deviation from HWE after Bonferroni correction.

Locus Name/ GenBank Accession no.		Motif	Forward Primer (5'-3')	Reverse Primer (5'-3')	*Ta*	Overall			North Sea				Irish Sea			
						N	*Na*	Range	*Na*	*Ho*	*He*	*p*	*Na*	*Ho*	*He*	*p*
DAC1-35	EU982372	(AC)40	GAAGTCTCCAGGAACGACTACA	TCAAGAACACAGACGTCAGGA	60	47	26	302-372	22	0.913	0.932	0.051	22	0.875	0.933	0.065
DAC1-55	EU982373	(AC)20	AAAGTGGGGATTGAGGAAGG	ACACCACACACCCACACAAT	60	48	10	242-268	9	0.792	0.745	0.558	8	0.667	0.802	0.07
DAC1-6	EU982374	(AC)38	GTCAGAACCACCCCCACA	TGAGACAGTTTGACCCTGATTTT	55	45	42	144-346	28	0.917	0.946	0.04	27	0.905	0.955	0.084
DAC1-90	EU982375	(AC)24	TGGCTCCTATCAAATACATA	CTCTGTTTCTTTCAGGACTC	60	48	21	102-146	18	0.958	0.918	0.241	19	0.917	0.925	0.185
DAC2-15	EU982376	(AC)24	CTCAGAGATGCCCAGAGGTC	GACAAGAACGCACGCACAC	60	48	11	174-216	8	0.625	0.681	0.072	8	0.708	0.723	0.313
DAC2-28	EU982377	(AC)10C(AC)5	GTGTTTCCGCTTGGCTTG	GCCTGGCAGACACCTACACT	60	48	12	110-150	10	0.875	0.807	0.715	9	0.875	0.767	0.283
DAC2-36	EU982378	(AC)18	GTTTGTTGCTCAGGTGCAGA	TGGGGAAGACACGTGTAAGA	52	47	34	215-327	21	0.913	0.791	1	23	0.958	0.894	0.933
DAC2-37	EU982379	(AC)11	GGTATGTGCTTTGCCCTCAG	TGTTTGGTTGTCCGTTATGG	58	48	4	240-248	3	0.292	0.254	1	3	0.083	0.081	1
DAC2-82	EU982380	(AC)50	ATGAAGCCTGTGTGCCTTTC	TTATGACCCTGGTTCCCTCA	55	45	34	335-434	24	0.913	0.946	0.031	28	0.909	0.953	0.239
DAC3-12	EU982381	(AC)15GC(AC)12	CTGCTTGTTTTGGTGACACA	TAGGCGTGTGTGCATATGTT	55	47	18	103-141	15	1	0.903	0.902	14	0.75	0.897	0.029
DAC3-14	EU982382	(AC)12	CTGTCAACTCGACTCTGGAGGA	GCAAGAACACACATATTCAGCTACA	60	48	8	160-174	6	0.542	0.726	0.006	6	0.5	0.697	0.043
DAC3-86	EU982383	(AC)15	GACCCCTCATGTGACTCCAG	CCTCTGAGGGCCCTTGTC	55	48	8	221-241	6	0.458	0.55	0.318	6	0.458	0.418	0.763
DAC4-20	EU982384	(AC)30	GTTTCCACGCTGCCTTCTT	TTCATCAATTTAACATAAAAAGAGAGA	55	45	30	123-181	25	0.917	0.932	0.633	22	0.955	0.944	0.698
DAC4-34	EU982385	(AC)15	TCCGGAGAGGTGAGGAGTTA	CATCGAATGAAAATGGAGGAG	55	46	31	179-241	21	0.591	0.93	**<0.001**	22	0.625	0.939	**<0.001**
DAC4-40	EU982386	(AC)21	TAGATAATGGGGCCCACAGG	TTAGCCGTTGTGGTTGACAG	60	47	20	320-365	15	0.957	0.872	0.795	17	0.875	0.881	0.26
DAC5-21	EU982387	(AC)11(AGACAC)5	AAATGTGACGTAGGTTAGGTTTCTG	CGAAGGCAGCTTTCTTCTCT	58	48	26	96-160	23	0.875	0.944	0.014	20	0.958	0.928	0.951
DAC5-5	EU982388	(AC)11AT(AC)6	TGCTTGAAGGCATTGTTGAC	CGTAGCTGCCTCTGAGTATTTG	60	48	5	117-127	5	0.25	0.264	0.295	3	0.208	0.223	0.113
DAC5-70	EU982389	(AC)13	CAGACATGTTTGTGTTTTCTCTCTG	AGGCACGAAAGCATGAATGA	58	48	28	112-188	14	0.75	0.875	**<0.001**	24	0.75	0.942	0.004
DAC5-77	EU982390	(AC)10	TCAATGGGGCAAAAGACAAT	CTTTCAATCGTGCATTCTTCA	60	48	8	104-122	6	0.333	0.33	0.419	7	0.5	0.495	0.03
DAC5-78	EU982391	(AC)18	AGGAATGAATCGTCCTGTGG	CAAACCACCAGGGGAATAAA	55	47	36	100-186	23	0.87	0.93	0.053	27	0.875	0.952	0.011
DAG1-14	EU982392	(AC)13	AAGGGATGATTGCACACACA	TGCAAAGGTTTGTTGAAGAACT	52	48	8	175-193	6	0.71	0.62	0.95	7	0.38	0.45	0.03
DAG2-15	EU982393	(AG)16	GACATGGCATCAGCTCTTGA	TCCCACAAGTAAAAGAAATTCCA	52	46	13	145-169	10	0.61	0.88	**<0.001**	12	0.57	0.88	**<0.001**
DAG2-22	EU982394	(AG)17	CGTTTACATGTGGTATCTGTCTG	AGATGGACAGATAGATGGATTGA	55	48	20	122-166	15	0.667	0.892	**<0.001**	14	0.458	0.883	**<0.001**
DAG2-90	EU982395	(AG)11AT(AG)5	AGGCAAGGATTTGGAAGGTT	TCACCCCTTAATCTGGAATTG	60	48	14	158-186	14	0.875	0.897	0.724	12	0.958	0.88	0.824
DAG4-64	EU982396	(AG)5GGG(AG)16	TGCACGTTGTGTGTCTCTCTC	GGGAAAAAGGAGGGGAAATA	60	48	23	143-191	20	1	0.933	0.741	18	0.958	0.919	0.539
DAG4-91	EU982397	(AG)24CG(AG)10	CTGCCGATGAAGGAGTTTTC	TGTGTGGTAGCAGACAGTGGA	60	45	30	209-345	23	0.583	0.92	**<0.001**	19	0.524	0.931	**<0.001**
DAG5-12	EU982398	(ATCT)23	CCCCAATTCATTATCTATGAACG	CCGGCAATCCAGGTTACTTA	55	47	21	132-256	18	0.958	0.887	0.229	13	0.957	0.843	0.78
DAG5-17	EU982399	(AG)29	ACCTGTCTGCAGGAAGAGGA	TCTGATGTGCTGCTGTTTCC	60	47	31	170-242	23	1	0.928	0.799	25	1	0.946	0.707
DAG5-45	EU982400	(AGAT)39	AAATAAGACTGGAATAAATATGCAC	AATATACCGGCTGCTATGAC	55	47	25	207-275	19	1	0.924	0.473	21	0.958	0.926	0.634
DAG5-88	EU982401	(AG)10AA(AG)9	TTTTCCCGAAAGTCCCTCTT	AGCCGGGATTTCATTATTCC	58	48	2	187-189	2	0.688	0.313	0.357	2	0.75	0.25	1

Table 2.2: Cross-species amplification of dab, *Limanda limanda*, microsatellites

Results of cross-species amplification of 30 microsatellite loci developed from dab on other European flatfish species (n=1/spp). Amplification conditions identical to those described for dab. Presence of microsatellite-like products is indicated with allele sizes, absence with a (-). Ll = *Limanda limanda*; Pf = *Platichthys flesus*; Pp = *Pleuronectes platessa*; Lw =*Lepidorhombus whiffiagonis*; Hp = *Hippoglossoides platessoides*; Sr = *Scophthalmus rhombus* ; Pm = *Psetta maxima*; Mv = *Microchirus variegatus*; Mk = *Microstomus kitt* ; Ss = *Solea solea*; Pl = *Pegusa lascaris*; Bl = *Buglossidium luteum*.

Locus	Ll	Pf	Pp	Lw	Hp	Sr	Pm	Mv	Mk	Ss	Pl	Bl
DAC1-35	342/356	-	-	-	-	-	-	-	-	-	-	-
DAC1-55	254/258	-	-	-	-	-	-	-	-	-	-	-
DAC1-6	249/249	144/170	-	144/144	196/196	158/160	162/162	152/154	-	-	-	185/189
DAC1-90	122/132	-	-	-	-	-	-	-	-	-	-	-
DAC2-15	202/204	-	-	-	-	-	-	-	-	-	-	-
DAC2-28	130/130	118/118	118/130	-	134/134	116/118	116/118	-	-	118/128	-	-
DAC2-36	231/271	-	-	-	-	-	-	-	-	-	-	-
DAC2-37	244/248	240/240	240/240	-	242/242	238/238	232/232	-	-	-	-	-
DAC2-82	349/437	345/361	-	-	-	-	359/389	-	-	-	-	-
DAC3-12	117/125	103/109	107/125	101/109	113/113	99/99	89/125	85/93	-	93/101	-	-
DAC3-14	164/172	-	-	-	-	-	-	-	-	-	-	-
DAC3-86	223/229	-	-	-	-	-	-	-	-	-	-	-
DAC4-20	136/160	-	-	-	160/166	-	-	-	-	-	-	-
DAC4-34	235/235	-	-	-	239/239	-	-	-	-	203/203	-	-
DAC4-40	334/351	-	324/324	-	-	-	-	-	-	-	-	-
DAC5-21	109/119	-	-	-	126/126	-	134/134	-	-	-	-	-
DAC5-5	121/121	119/121	115/121	-	111/121	121/121	121/121	-	-	109/115	121/121	121/121
DAC5-70	146/154	-	-	-	-	-	-	-	-	-	-	-
DAC5-77	108/112	-	106/106	102/138	102/138	102/138	102/138	-	-	102/106	102/106	102/138
DAC5-78	131/153	-	161/161	-	-	101/103	114/124	-	-	133/139	-	-
DAG1-14	183/183	-	-	-	181/187	-	181/181	-	178/181	181/209	-	-
DAG2-15	155/155	-	-	-	-	-	-	-	-	-	-	-
DAG2-22	138/138	108/156	-	124/138	138/138	-	156/156	-	-	136/264	-	122/134
DAG2-90	162/176	178/186	-	-	170/182	156/176	152/156	-	-	-	-	-
DAG4-64	147/159	-	-	-	-	-	-	-	-	-	-	-
DAG4-91	260/260	-	-	-	-	-	-	-	-	-	-	-
DAG5-12	223/267	241/341	-	-	137/145	-	203/211	-	-	-	-	-
DAG5-17	200/208	-	-	-	168/182	-	174/212	-	-	-	-	-
DAG4-64	147/159	-	-	-	-	-	-	-	-	-	-	-
DAG4-91	260/260	-	-	-	-	-	-	-	-	-	-	-

Chapter 3: Isolation and characterisation of 28 microsatellite markers for European flounder (*Platichthys flesus* L.).

3.1 Abstract:

European Flounder (*Platichthys flesus* L.) are used in ecotoxicological studies to provide detailed information on the effects of pollution on individual fish. Data on population and evolutionary level effects are, however, limited. Here, the isolation and characterisation of 28 novel species specific microsatellite loci are presented. The number of alleles ranged from 8 to 38, and observed heterozygosity from 0.542 to 1.

3.2 Introduction:

The European flounder, *Platichthys flesus* L., is used as a model species in ecotoxicology to examine the ecosystem impacts of anthropogenic pollution. For example, recent studies have examined the effects of toxicants on DNA disruption, gene expression, transcriptomics and tumorigenesis (Stentiford *et al.*, 2003; Marchand *et al.*, 2006; Williams *et al.*, 2007). Although some microsatellite markers have been used on European flounder (Casas *et al.*, 2005; Hemmer-Hansen *et al.*, 2007b), only six non-EST-linked species-specific loci have been published to date. Increasing this number will extend the range of applications, such as population assignment, localised impacts of toxicant exposure, and ultimately contribute towards the production of linkage maps. Here, the development and characterisation of 28 novel microsatellite markers developed from European flounder is described.

3.3 Methods, Results and Discussion:

European flounder (n=2) were collected from the Irish Sea (53°18'56.45"N; 3°53'59.64"W) and fin clips stored in 100% ethanol until processing. A microsatellite enriched genomic library was constructed following the protocol of Tysklind *et al.* (2009b). From this library, 1344 colonies were screened for microsatellites by PCR amplification with M13 Forward primer and a mixture of non-biotinylated microsatellite probes. 196 positive amplicons were sequenced (Macrogen Inc., Korea), edited, analyzed and checked for duplicates using BIOEDIT (Hall, 1999). Enrichment efficiency was high at 87.8%. Primers were designed either side of 47

putative microsatellites using PRIMER3 (Rozen & Skaletsky, 2000), and tested for successful amplification at several annealing temperatures on 3% TBE agarose gels. M13-tailed forward primers were then ordered (Schuelke, 2000). Nested PCRs with tailed forward primer, reverse primer, and FAM™-labelled M13-tail oligos were used for genotyping. PCR cocktails of 10 µl final volume contained around 20 ng of DNA, 1x GoTaq® Flexi buffer (Promega), 1.5 mM $MgCl_2$, 125 µM dNTP, 0.1 µM Forward-tailed primer, 0.5 µM of Reverse primer, 0.5 µM of FAM™-labelled M13-tails, and 0.5 U GoTaq® DNA polymerase (Promega). PCRs were carried on a BioRad Tetrad2® Peltier Thermal Cyclers and the thermocycling programmes were as follows: an initial denaturation phase of 3 min at 95°C, followed by 13 cycles of 30s at 95°C, 45s at 55°C, 60s at 72°C, then 31 cycles of 30s at 94°C, 45s at 50°C, 60s at 72°C, and finishing with a 30min extension phase at 72°C.

Two samples of 24 individuals collected from the Mersey Estuary (53°19'28.70"N; 2°54'39.74"W) and the Tyne estuary (54°59'10.21"N; 1°27'49.47"W) were genotyped on an ABI 3130xl Genetic Analyzer (Applied Biosystems) with an internal size standard (GeneScan™ LIZ-600®). 28 primer pairs produced polymorphic bands at the expected sizes. Allele sizes were scored with GeneMapper® Software 4.0. Genotypes were analyzed with GENETIX (Belkhir *et al.*, 1996-2004) and GENEPOP V4.0 (Rousset, 2008), where polymorphism varied from 8 to 38 alleles with an average of 16 alleles per locus per population (Table 3.1). Observed heterozygosity ranged from 0.542 to 1, and only one marker (FLAC4-60) significantly deviated from Hardy Weinberg Expectations in the Tyne sample after Bonferroni correction. FLAC2-18 was similar to *Sciaenops ocellatus* microsatellite Scoc76 (GenBank: EU727070.1). Markers FLAC1-32 and FLAG2-76 appear to be in linkage disequilibrium in both populations and overall, suggesting that both markers are gametically linked. Cross-species amplification on eleven European flatfish species was tested using the same parameters as for flounder (Table 3.2).

These novel nuclear markers will contribute to the library of available markers and will provide tools to enhance our understanding of the genetic structure of flounder throughout its range, and increase the power of detecting signals of evolutionary processes imposed on populations under contaminant and other environmental stress.

Table 3.1: Characteristics of 28 microsatellite loci for flounder, *Platichthys flesus*

Characterization of 28 microsatellite loci isolated from *Platichthys flesus* in two estuarine populations. Motif = repeat sequence of the isolated clone; *Ta* = Annealing temperature; *N*= number of individuals successfully amplified (out of 48). *Na* = Number of alleles; Range = allele size range; *Ho* = Observed Heterozygosity; *He* = Expected Heterozygosity; *p* = associated probability value of conformation with Hardy-Weinberg Equilibrium (HWE). **Bold** p values indicate significant deviation from HWE after Bonferoni correction.

Locus Name/ GenBank Accession no.		Motif	Forward	Reverse	Ta	Overall			Mersey Estuary				Tyne Estuary			
						N	Na	Range	*Na*	*Ho*	*He*	*p*	*Na*	*Ho*	*He*	*p*
FLAC1-32	FJ360492	(AC)19	GAGAAACCTCCCACAGGTGA	GGGAAATAACGCTGATACGAA	55	48	13	156-184	9	0.833	0.829	0.381	11	0.958	0.881	0.070
FLAC1-47	FJ360493	(AC)7+(AC)16	ACGCACATACGAAGCCGTAG	ATTTCTGCCCAGGGATTACC	55	47	10	242-270	7	0.542	0.680	0.044	8	0.652	0.700	0.500
FLAC2-16	FJ360494	(AC)15	CCATCTACCCCCAGAAAACA	CTCCGGGGGAAACTTAAGAG	55	48	20	156-202	17	0.917	0.869	0.431	16	0.833	0.904	0.022
FLAC2-18	FJ360495	(AC)20	AAAAGCCAATGACCTGTTGC	GCATGCCAGTGAGAACAATG	55	48	20	216-266	16	0.958	0.910	0.041	18	0.833	0.901	0.074
FLAC2-46	FJ360496	(AC)16	TGCATGGTTTTAAATACGACAA	GAACAGCAGCCTTTGTTTTTG	55	45	16	167-209	13	0.875	0.868	0.316	14	0.810	0.892	0.129
FLAC3-19	FJ360497	(AC)5AC(AC)17	GAAACCAGGGGGCTCTTTAG	TCTTAGATCCCCCGAATGTG	55	48	11	221-279	8	0.792	0.813	0.579	11	0.833	0.858	0.537
FLAC4-60	FJ360498	(AC)22TC(AC)5	CGCCTCTTGACACACAGAGA	AGCCCTTCTGTTCGAGTTCA	55	48	18	179-257	13	0.833	0.859	0.877	15	0.792	0.893	**0.001**
FLAC4-67	FJ360499	(AC)27	ACCTGGACCAAACACACACA	CCCCACCATGTCAGAACTTA	55	48	27	175-296	19	1.000	0.914	0.613	22	0.875	0.931	0.206
FLAC4-69	FJ360500	(AC)5+(AC)14	GGGAGAACCTGTCAAACCTG	AACGCAACACGCATACAAAC	55	47	23	228-300	20	0.875	0.934	0.161	19	1.000	0.919	0.995
FLAC4-7	FJ360501	(AC)5AT(AC)3GC(AC)23	CGATGGCCTTCTCTTTTTCA	TTTTGATGCCATTTGCATGT	55	47	12	169-197	12	0.913	0.837	0.969	10	0.792	0.806	0.470
FLAC4-81	FJ360502	(AC)16AG(AC)4	CAAGGCTCCTGTGACAGCTT	CCAGTAAGAGATCAAACACG	55	48	14	174-208	12	0.708	0.867	0.020	12	0.625	0.850	0.006
FLAG1-11	FJ360503	(AG)18	GCGAGAGAGGGAGAAAGAAAA	TTCCTATCTGGTTCAGTCCTTCTT	55	48	23	235-337	18	0.833	0.914	0.021	18	0.917	0.919	0.622
FLAG1-81	FJ360504	(AG)20GGAGGG(AG)8	AAAACCTGGTGGCATATGGT	GCTTTCATTTCCCAGGTCAG	55	47	20	196-236	17	1.000	0.899	0.688	14	0.833	0.892	0.068
FLAG2-76	FJ360505	(AG)6CG(AG)21	ACCTTTCGACCACCTGTCTG	TCTCATGTGCTGCTGTTTCC	55	43	25	156-226	17	0.800	0.918	0.146	21	0.870	0.928	0.034
FLAG3-55	FJ360506	(AG)16	GCCAGCTTCATGACACACAC	GCCGATGGCATGTAGAGAAT	55	48	13	228-266	8	0.833	0.686	0.791	11	0.833	0.797	0.064
FLAG3-73	FJ360507	(AG)16	TGATGGACAGCTTCAGCATC	GAACACCATCAGGTATCTTCATCA	55	48	23	257-325	16	0.917	0.903	0.115	18	0.875	0.909	0.025
FLAG4-25	FJ360508	(AC)15A(AC)19	CGGGGTCACAGTTTAACACA	TGTTCATGTGGTTGCATTTG	55	46	37	120-314	25	0.957	0.947	0.670	22	0.870	0.908	0.039
FLAG4-65	FJ360509	(AGAT)18+(AG)11	TGTGTGTGAGTGTATGTTTACTTGG	CAGTTGCGCAAGCTAATGTC	55	48	24	232-302	19	0.833	0.926	0.024	22	0.958	0.941	0.506
FLAG4-71	FJ360510	(AG)17AA(AG)5	TGAAAAGGGATAAGAGGGAGA	TTCTAGCTGGACTCAAGGGTAA	55	46	30	183-247	24	0.917	0.947	0.290	20	0.955	0.934	0.546
FLAG5-83	FJ360511	(AGAT)16	CCAGTGCAGAGGAGTTTTCAG	TTGAGTTCACTCCTGCACCA	55	41	38	200-340	28	1.000	0.957	1.000	24	0.895	0.950	0.076
FLAG5-87	FJ360512	(AG)9GG(AG)16	TCTTCTCGCTGCATGAACAC	CCGTTTCCTTTGTCCAACAT	55	48	23	159-263	18	1.000	0.916	0.818	20	1.000	0.917	0.017
FLAG6-14	FJ360513	(AG)18AT(AG)8	CAGTAGCAGGGTGTTTTTCCTT	CCAGTAATGACCAAACCCAAA	55	44	31	226-364	21	0.857	0.925	0.026	21	0.870	0.936	0.009
FLAG6-77	FJ360514	(AGAT)33(AT)3(AGAT)9	TAAGATAGATAGCTGCATTG	CATGTTTCTTTCACAAATTA	55	48	29	136-274	23	0.833	0.944	0.043	22	0.958	0.940	0.726
FLAG7-17	FJ360515	(AG)18	GCAACGAGCTGCTAATTAAGG	CGACCAACACAAACCACTTG	55	48	8	170-188	6	0.917	0.759	0.251	7	0.875	0.747	0.864
FLAG8-19	FJ360516	(AG)11GG(AG)13	ACCTCGGCCAGCACTTAATA	AAAGGGGGCAGATGATTAGG	55	48	20	165-209	16	0.875	0.901	0.555	18	0.917	0.932	0.344
FLAG8-37	FJ360517	(GT)10G(AG)8CG(AG)30	GAACTCCTGTCCTGCTGCTC	CCGTCATCGCTCTCTGAGG	55	47	24	213-275	16	0.826	0.872	0.299	23	0.875	0.930	0.248
FLAG8-41	FJ360518	(AG)19	AAATCCAGATGCAGGTCACA	GAGGCTCTGGCTGTTTGTTC	55	48	13	221-255	11	0.917	0.783	0.068	12	0.792	0.832	0.243
FLAG8-89	FJ360519	(AC)12+(AG)23	CCCATACAGACAGCTGGTGA	TTTCCCACGATGGAGGAG	55	47	23	173-223	17	1.000	0.919	0.909	20	0.875	0.915	0.402

Table 3.2: Cross-species amplification of flounder, *Platichthys flesus*, microsatellites

Results of cross-species amplification of 28 microsatellite loci developed from European flounder on other European flatfish species (*n* = 1/spp.). Amplification conditions identical to those described for European flounder. Presence of microsatellite-like products is indicated with allele sizes, absence with a '—'. Pf = *Platichthys flesus*; Ll = *Limanda limanda*; Pp = *Pleuronectes platessa*; Hp = *Hippoglossoides platessoides*; Lw = *Lepidorhombus whiffiagonis*; Sr = *Scophthalmus rhombus*; Pm = *Psetta maxima*; Bl = *Buglossidium luteum*; Ss = *Solea solea*; Pl = *Pegusa lascaris*; Mv = *Microchirus variegatus*; Mk = *Microstomus kitt.*

Locus	Pf	Ll	Pp	Hp	Lw	Sr	Pm	Bl	Ss	Pl	Mv	Mk
FLAC1–32	176/176	168/170	166/170	166/166	—	166/178	—	—	160/162	—	154/166	—
FLAC1–47	262/262	—	—	—	—	—	—	—	—	—	—	—
FLAC2–16	178/196	170/172	192/196	164/172	160/160	164/166	164/166	158/192	176/180	158/258	—	—
FLAC2–18	242/246	256/260	250/256	226/226	—	—	246/248	—	266/270	280/280	302/308	296/296
FLAC2–46	187/205	179/191	181/209	—	—	—	—	—	195/207	—	—	—
FLAC3–19	221/273	—	249/249	259/263	—	249/265	253/265	—	251/251	267/267	237/263	267/279
FLAC4–60	179/183	—	—	—	—	—	—	—	—	—	—	—
FLAC4–67	192/199	—	—	—	—	—	—	—	—	—	—	—
FLAC4–69	276/285	—	290/292	—	—	—	—	254/264	242/242	—	246/280	
FLAC4–7	187/189	—	193/193	—	—	—	—	—	—	—	—	—
FLAC4–81	174/192	—	192/198	176/188	—	—	—	—	—	—	186/194	182/192
FLAG1–11	249/249	—	—	—	—	—	—	—	—	—	—	—
FLAG1–81	204/210	—	—	222/222	—	184/190	—	—	—	—	—	—
FLAG2–76	170/170	—	—	—	—	—	—	—	—	—	—	—
FLAG3–55	256/258	—	230/236	—	—	—	—	—	228/252	—	—	—
FLAG3–73	277/295	—	—	—	—	—	—	—	—	—	—	—
FLAG4–25	254/284	—	—	—	—	—	—	—	—	—	—	—
FLAG4–65	240/286	—	180/188	—	264/264	—	160/198	158/194	182/224	166/206	154/190	164/202
FLAG4–71	197/211	—	185/195	—	—	—	187/205	—	—	211/227	—	—
FLAG5–83	314/314	—	—	—	—	—	—	—	—	—	—	—
FLAG5–87	167/177	—	—	—	—	—	—	—	—	189/189	—	187/199
FLAG6–14	304/324	—	273/277	—	256/266	—	228/244	—	—	—	264/282	—
FLAG6–77	196/220	—	148/194	—	—	—	124/160	—	168/176	—	120/156	126/164
FLAG7–17	178/180	168/168	172/180	178/190	172/174	176/176	176/176	174/182	176/180	—	—	168/174
FLAG8–19	171/183	190/205	187/195	—	—	167/199	—	193/197	185/190	181/185	167/185	179/183
FLAG8–37	235/237	—	—	202/202	—	—	—	—	—	—	—	—
FLAG8–41	239/239	—	237/237	—	—	—	—	—	241/248	221/221	245/248	253/267
FLAG8–89	195/209	—	207/225	195/195	—	221/221	219/219	—	171/183	185/197	181/189	—

Chapter 4: Maximising the cost to benefit ratio of microsatellite genotyping by multiplex tailing

4.1 Abstract:

Microsatellite markers are used in ecological studies to answer increasingly elaborate questions. The potential to reach accurate conclusions is dependent on the number of markers employed, and thus, there is a need to develop cheaper and more effective ways of genotyping non-model organisms. Here, a set of oligonucleotides (tails) with similar annealing temperatures, GC content, and no self-priming or cross-priming properties are suggested and tested in multiplex PCRs. The economic, technical, and logistic implications of the general use of these tails are also discussed.

4.2 Introduction:

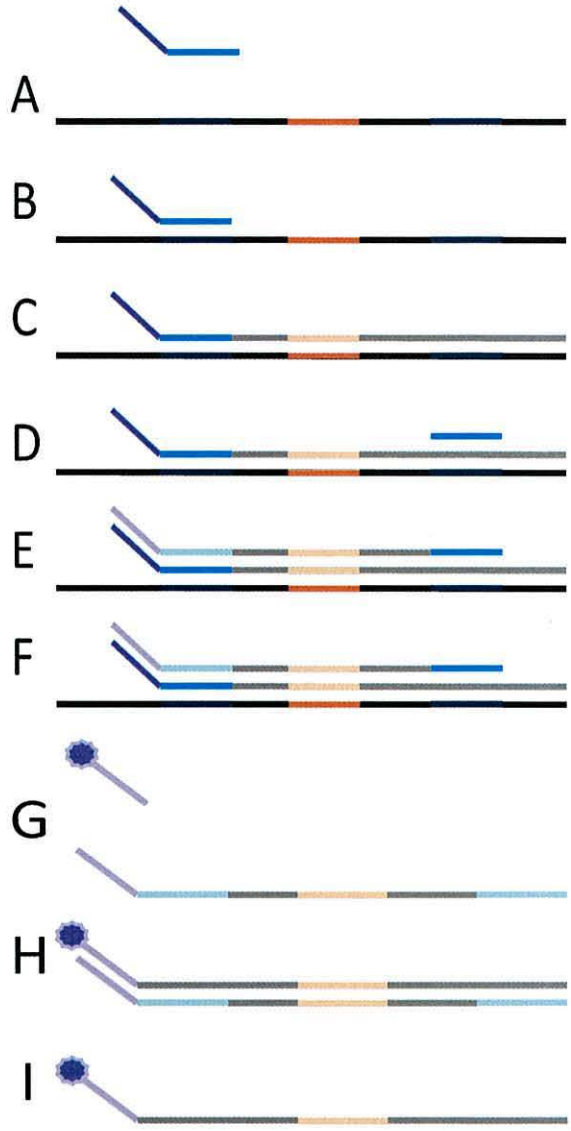
Since the mid 1990's, microsatellites have been the marker of choice for parentage analysis, assessment of neutral variation and population structure in natural populations, and genome mapping (Jarne & Lagoda, 1996; Li *et al.*, 2002; Chistiakov *et al.*, 2006; Oliveira *et al.*, 2006). However, their high initial development cost has often restricted their use in species with no existing genomic resources. The cost of development in novel species may be overcome by using microsatellites developed from closely related species (Primmer *et al.*, 1996). However, cross-amplifying microsatellites from closely related species may lead to problems with reduced polymorphism and non-specific amplification due to ascertainment bias (Oliveira *et al.*, 2006). Ascertainment bias is the result of the several stages of selection for polymorphism during the microsatellite development in the target species that lead to reliable and polymorphic loci in the target species but not necessarily in other taxa (Goldstein & Pollock, 1997; Brandström & Ellegren, 2009). Therefore, many researchers prefer to develop species-specific loci despite the costs. Fortunately, recent technical advances have simultaneously reduced the cost and increased the yield of polymorphic microsatellite loci (Zane *et al.*, 2002).

After the initial stages of the microsatellite development process (i.e. partial library construction, probing and sequencing), a relatively large number of microsatellite sequences may be produced, for which designing and testing primers is simple and relatively cheap. Oligonucleotide primers are designed that are complementary to the flanking sequence on

the 5′ and 3′ sides of the microsatellite sequence, and the microsatellite is then amplified using the polymerase chain reaction (PCR). After initial amplification to verify the utility of the primers and the approximate size of the amplified fragment, the forward primers of successfully amplifying primer pairs are labelled with a dye that fluoresces under excitation by a laser, enabling the size of the PCR product to be accurately measured in an automated sequencer. Modern automated sequencers can measure fluorescence at different wavelengths allowing multiple dyes to be used to label PCR primers. This property of modern sequencers allows the simultaneous measurement of multiple PCR fragments of the same size (if they are labelled with different dyes), and also, the measurement of PCR products of different sizes labelled with the same dye. Thus, both size and dye-based discrimination is possible, and combining several labelled primers is possible to reduce the time and costs associated with microsatellite genotyping (Neff *et al.*, 2000; Schoske *et al.*, 2003). However, despite the savings that are possible by multiplexing primers, the initial costs of purchasing labelled primers can be prohibitive, thus constraining the number of microsatellites tested, and ultimately the total number of markers used in a genotyping project.

New analysis methods and programs are tackling ever more complex questions regarding population structure, migration rates, effective population size (N_e), demographic changes and relatedness (Excoffier & Heckel, 2006), that require increasing numbers of loci to provide robust answers. Furthermore, the power of any analysis is generally improved with a higher number of loci (Kalinowski, 2002b; Medina *et al.*, 2006; Ryman & Palm, 2006). Hence there is an increasing demand for fast, simple, and economic hi-throughput techniques for analysing large numbers of loci. Schuelke (2000) described a nested PCR in which a universal fluorescent oligonucleotide (M13 tail) could be used to label any PCR product, a technique that is frequently used during microsatellite development (Canino *et al.*, 2005; Johansson *et al.*, 2008; Lallias *et al.*, 2009; Nikolic *et al.*, 2009; Tysklind *et al.*, 2009b; 2009a). A *tail* is essentially an oligonucleotide extension of known sequence which is added to the 5′ end of the forward primer upon design (i.e. tail-primer). In order to work effectively, the tail should not interfere with the priming properties of the primer, and thus the PCR can proceed as normal, with the only exception that during the PCR the DNA polymerase incorporates the tail sequence into the amplicon. The advantage of adding such tail to the primer is that another oligonucleotide, a fluorescently *labelled-oligo* of exactly the same sequence as the tail, can be added to the PCR reaction. Thus, the labelled-oligo will anneal with the area of the amplicon composed by the original tail, and after enough cycles of amplification, obtain a labelled amplicon (Figure 4.1 A-I)

Figure 4.1: Tailing process in a PCR



(A) a sequence of DNA is represented (Black). The target to be amplified is represented in orange. A forward tailed-primer in blue (forward) and purple (tail) are also drawn.

(B) The forward tailed-primer anneal to the binding site in the DNA sequence

(C) The DNA polymerase extends the sequence from the priming site, replicating the target

(D) The reverse primer anneals to the reverse binding site

(E) The DNA polymerase extends the sequence from the reverse binding site.

(F) The process is repeated several times

(G) Once all of the tailed-primer has been incorporated into amplicons, the labelled-oligo is allowed into the reaction

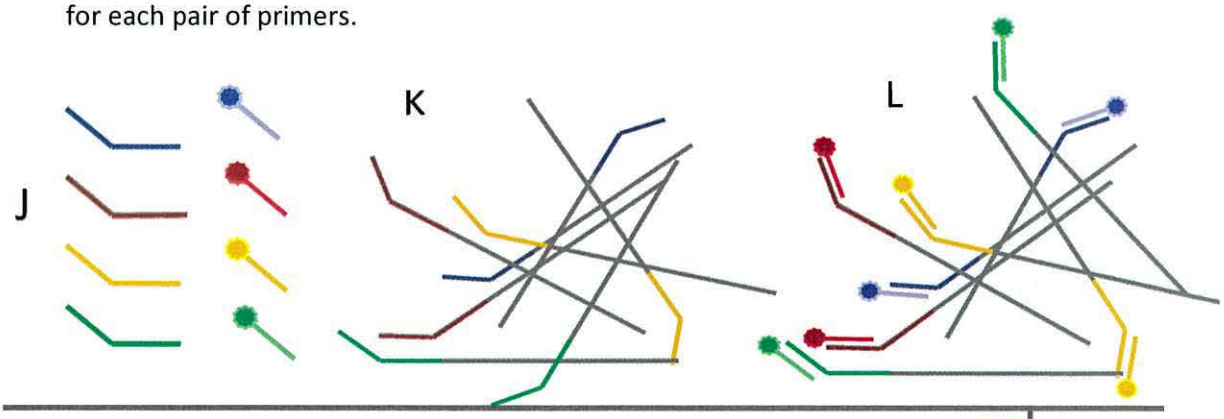
(H) The fluorescent-oligo binds to the tailed section of the amplicon

(I) A fluorescently labelled amplicon is produced

(J) Four different forward tail-primer with different tails (different colours) are designed, with correspondingly different labelled-oligos.

(K) The PCR reaction is started, and the forward tail-primers are incorporated into many different amplicons.

(L) The labelled-oligos only bind to specific tail-primers thus creating label-specific amplicons for each pair of primers.



Previously, several PCR products have been labelled with the same labelled tail within a single reaction (Oetting *et al.*, 1995). However, the technique is not cost-effective at the genotyping stage as the other detection wavelengths of modern DNA sequencing platforms are not utilised. Guo & Milewicz (2003) successfully multiplexed microsatellite primers with two different tails labelled with two different dyes, improving the efficiency of the tailing protocol, but still not exploiting other wavelengths. Missiaggia & Grattapaglia (2006) addressed this issue by performing separate PCR reactions for each colour (with four M13 tails in different colours) and then blending them together before sequencing, a technique generally known as poolplexing (Meudt & Clarke, 2007). Although the latter approach makes full advantage of detection capability of the sequencer, it is reagent and time consuming and becomes unpractical for large genotyping projects.

Here, a cost-effective way of labelling primers in four different wave-lengths for multiplex microsatellite analysis of large sample sizes is described.

4.3 Methods, Results and Discussion:

4.3.1 Tailing:

The aim was to use a limited number of labelled-oligos to label several target amplicons with different dyes in a single reaction. The underlying principle was to allow the tailing process for each of the dyes employed to occur independently, and in parallel, within a single reaction (Figure 4.1 J-L). To achieve these conditions, the four tails must not interfere with each other (i.e. not bind to each other), and operate under the same conditions (i.e. annealing temperatures). Therefore three additional oligonucleotide tails that fulfilled the following conditions were identified: the same number of bases (18bp), similar GC content (50-55%), similar annealing temperature (53-55°C), and no self-priming or cross-priming between tails, which was tested theoretically using FASTPCR (Kalendar, 2009). Of the 19 universal bacterial sequences considered, three fulfilled the above conditions and were deemed suitable to work in multiplex with M13for tail: M13rev; Bhg-r; and +19bs (Table 4.1).

Microsatellite forward primers with M13for, M13rev, bhg-r, and +19bs tails at the 5' end were ordered. M13for, M13rev, bhg-r and +19bs oligos were ordered, each labelled with a different fluorescent dye (FAM®, PET®, NED®, VIC®). A subset of normally labelled forward microsatellite primers were ordered for amplicon comparison with tailed products, and all primers were normalized to 50µM stock.

Table 4.1: Oligonucleotide multiplexing tails:

Four oligonucleotide sequences (in lower case) suitable for use as labelled tails and compatible in multiplexing, followed by two examples of forward primers (*DAC2-15for* and *DAC2-28for* in UPPER case) and their tailed forms (*DAC2-15-P* and *DAC2-28-V*).

Name	Sequence	Label
M13for	tgtaaaacgacggccagt	FAM®
M13rev	gccgctctagaactagtg	VIC®
Bhg-r	tagaaggcacagtcgagg	NED®
+19bs	gcaggaaacagctatgac	PET®
Examples	**Sequence**	
DAC2-15for	CTCAGAGATGCCCAGAGGTC	
DAC2-15-P	gcaggaaacagctatgacCTCAGAGATGCCCAGAGGTC	
DAC2-28for	GTGTTTCCGCTTGGCTTG	
DAC2-28-V	gccgctctagaactagtgGTGTTTCCGCTTGGCTTG	

4.3.2 Multiplexing:

Dab microsatellite primers (Tysklind *et al.*, 2009b) were first grouped into potential multiplex combinations with MULTIPLX (Kaplinski *et al.*, 2005) which evaluates all possible primer pair compatibility issues (primer-primer or primer-product alignments, and differences in melting temperatures) and produces theoretical multiplex combinations. The stringency of the theoretical reaction can be modified, and primer combinations were evaluated at both normal and high stringency. Combinations of two or three primer pairs that consistently clustered together were chosen as starting nuclei for multiplex combinations. If amplification was successful, then additional primer pairs, checked for microsatellite allele size range, colour and primer compatibility in FASTPCR (Kalendar, 2009), were subsequently added to the multiplex. The process was repeated until no more primer pairs fitted the above conditions. Three multiplexes, containing 16 loci for dab, were successfully designed and tested (Table 4.2) and reactions were performed in 5µl PCR using the Qiagen® multiplex PRC kits with the following final concentration of primers: 0.2µM for reverse, 0.02µM for tail-forward, and 0.1µM labelled-oligo/loci in that colour (i.e. in a 12 loci mix, four loci in some colours = 0.4µm of each labelled tail) and around 20ng of DNA. The PCR reaction programme was modified from that in Hinten *et al.* (2007) and started with 15min at 95°C to activate the Hot-Start polymerase. Then 13 cycles of 30s at 94°C, 90s at 60°C, and 60s at 72°C, followed by 31 cycles of 30s at 94°C, 90s at 50°C, and 60s at 72°C, and finalised with a 30min extension phase at 60°C. The forward tail-primer anneals to the substrate DNA and is incorporated into the amplicon during the first 13 cycles, while the annealing temperature is too high for the labelled-oligos to anneal. In the following 31 cycles with a lower annealing temperature the labelled-oligos take over, incorporating the fluorescent label into the PCR amplicons. The different tails were initially tested singly and then in combinations into 4, 8 and 12 loci

multiplex. PCR products were resolved on an ABI 377 automated sequencer, and microsatellite allele sizes scored with GENEMAPPER software.

Multiplex	Fluorescent dye			
A	FAM®	VIC®	NED®	PET ®
	DAC5-5 114-126bp	DAC2-28 109-158bp	DAG2-90 159-177bp	DAC2-15 174-215bp
	DAG4-64 140-225bp			DAC1-55 236-270bp
	DAC1-35 288-346bp			
B	FAM®	VIC®	NED®	PET ®
	DAC5-77 107-113bp	DAC3-14 158-173bp	DAC4-40 305-343bp	
	DAG5-17 169-229bp			
C	FAM®	VIC®	NED®	PET ®
	DAC5-21 113-154bp	DAC1-90 109-142bp	DAC2-37 238-247bp	DAC5-70 89-135bp
		DAG5-88 180-192bp		

Table 4.2: Multiplex tables:

Description of three multiplex combinations (**A**, **B**, **C**) of microsatellite primers for single-reaction amplification in dab (*Limanda limanda*). Four fluorescent colours were employed (FAM®, VIC®, NED® and PET®), and several loci were sometimes amplified in each colour. Locus names and size ranges of amplified products are listed underneath the labelling dye employed for the locus.

4.3.3 Performance and Evaluation of multiplexed tailed primers:

The four oligonucleotide tails performed satisfactorily, both singularly and in multiplex, in over 60 microsatellite markers used in this study. Peak height and shape did not change significantly between tailed+(labelled-oligo), and labelled-only primers and were easily scored, the only difference being the size of the PCR product which was consistently 18bp longer. Although PCRs with up to 12 loci were tested successfully without reduction in electropherogram performance, problems with overlapping size ranges and incompatibility between primers posed considerable problems when searching for compatible groups of primer pairs, which resulted in multiplexes with relatively few loci (Table 4.2; Figure 4.2).

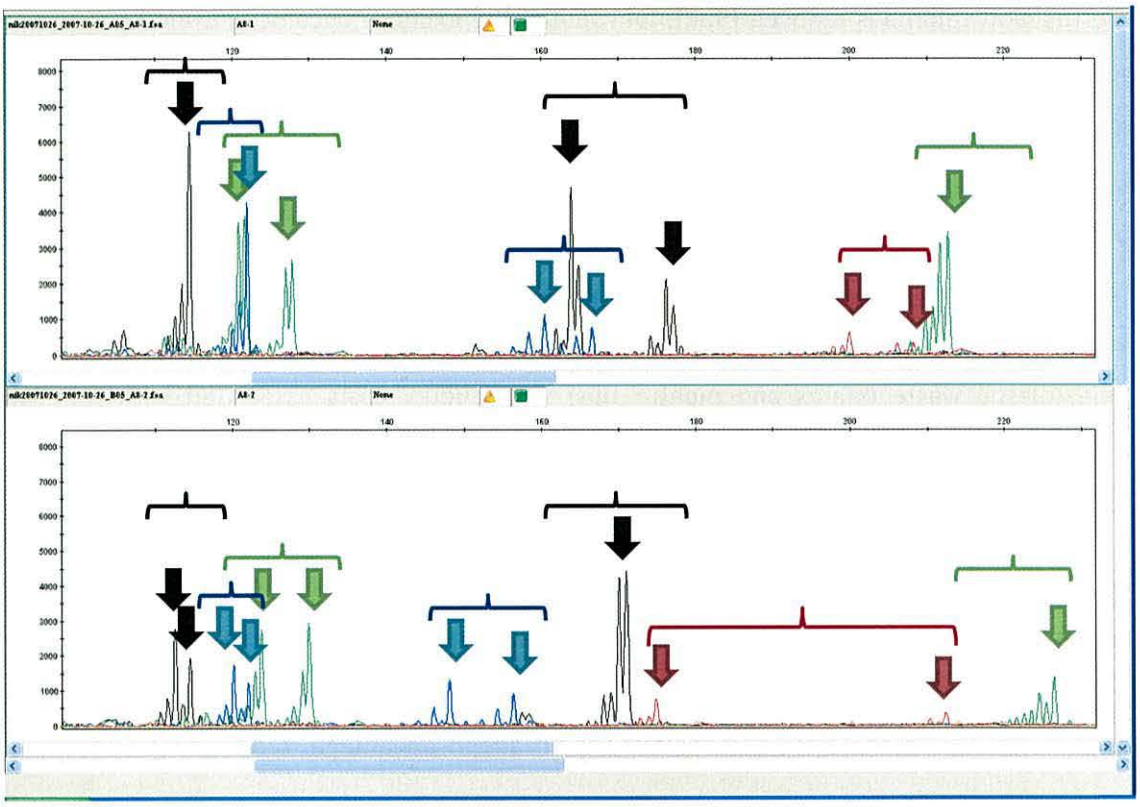


Figure 4.2: Multiplex electropherograms.

Multiplex electropherograms revealing the genotypes at seven loci of two individual dab. The alleles are indicated by arrows, and alleles belonging to the same locus are under the same bracket. Two loci are labelled with the *M13for* blue tail (FAM®), two loci with the *M13rev* green tail (VIC®), two loci with the *Bhg-r* yellow/black tail (NED®) and one with the *+19bs* tail in red (PET®).

The cost of designing tailed-primers (ca. £5.5 from a major oligonucleotide provider) is a fraction of the cost of fluorescently labelled primers (ca. £55 for 10,000 picomoles from Applied Biosystems). In the current study more than 130 primer pairs (plus around 30 loci in other organisms) have been tested at the sequencer level, which would have entailed a price of over £7,000 in labelled primers alone (130 x £55= £7,150), with many labelled primers being leftover from those loci not extensively employed (i.e. not included in multiplexes). Instead, all primers were labelled for under £1,200 by ordering all forward primers tailed (130 x £5.5 = £715) and four labelled tails (£35 for FAM® tail, and £135 for 300,000 picomoles of VIC®, NED® and PET® tails = 3 x £135 + 1 x £35 = £440). Note that labelled primers are considerably more economic in larger volumes (£55 for 10,000 picomoles compared to £135 for 300,000 picomoles).

Tailing microsatellite primers brings about considerable cost reductions that can only be exploited in the genotyping phase if used in multiplex. Purpose-built multiplexing PCR kits are

expensive (Qiagen: £1,197 for a 1000 x 50µl reactions), but very effective at multiplexing. If the reaction volume is reduced (5µl total volume) the expense becomes justified (£0.12 per reaction). In the dab genotyping study, 16 loci were genotyped in plates with just three multiplex reactions (plate cost = £1; reaction cost for 96 individuals: £0.12 x 96 x 3 + £1 x3 = £37.56), instead of 16 independent reactions with a cheaper PCR kit (£0.05 per 5µl reaction; plate reaction cost: £0.05 x 96 x 16 + £1 x16 = £92.80). The difference represents a 40% reduction in reaction cost alone, which would be even more pronounced as more loci and larger multiplexes are screened. Furthermore, there are considerable savings in technical time, plastic waste (plates and pipette tips), and energy costs associated with reduced number of reactions.

Additionally, the technique was tested on other fish, mollusc and reptilian species with equally impressive results suggesting they could be used on many other eukaryotes (Y. Surget-Groba; S. van Wijk; S. Pascoal; personal communication). These tails offer the advantage of being fully compatible with multiplexing in a single tube, overcoming the need to perform several PCRs for each colour. As an added advantage, redesigning primers, in order to change the size of the product or to avoid a polymorphic primer annealing sequence (null alleles), becomes more affordable.

Overall, the combination of tailing and multiplexing can substantially reduce the cost of genotyping projects, and in particular, molecular ecology laboratories, that commonly embark on many small and medium scale genotyping projects with many different species, will benefit the most: labelled tails can be purchased commonly, aliquoted, and used for several projects. Another added advantage of using common labelled tails is that larger number of users imply faster turnover of fluorescent oligonucleotides, which over time decrease in fluorescent strength.

Chapter 5: Development and testing of adaptive markers to evaluate signals of selection imposed by pollution in dab, *Limanda limanda.*

5.1 Abstract:

In the UK, dab is routinely used as an environmental bioindicator of pollutant exposure. However, pollutants can exert strong selective pressures on wild populations, and the potential effects of pollution from an evolutionary perspective have received less attention. Recent advances in genomics and bioinformatics have produced a rapid increase in the volume of DNA sequence data available, even for non-model organisms. Such data represent an invaluable resource for the development of genetic markers to study selection and adaptation processes. Here, the development and testing of 35 gene-linked microsatellites markers (EST-SSRs) derived from a library of differentially expressed genes after exposure to pollutants is reported. The markers are tested among putative healthy dab and dab exhibiting two common biomarker responses to pollution. No definitive evidence of strong and recent selective pressures at the analyzed genes is found, but suggestions for future research are made.

5.2 Introduction:

There is natural variation in the way that different individuals respond to environmental contaminants (Hawkins *et al.*, 1989; Weis & Weis, 1989; Depledge, 1990; Theodorakis & Shugart, 1998; Theodorakis *et al.*, 1999; Maes *et al.*, 2005), and this can be problematic when assessing the biological effects of pollutants (Depledge, 1990; 1996; Pulsford *et al.*, 1995; Belfiore & Anderson, 2001; Theodorakis, 2001; Meyer *et al.*, 2002; Kirby *et al.*, 2006). The evidence for genotype-specific sensitivity or resistance to disease development, especially tumour development, is accumulating in human medicine: certain polymorphisms at DNA repair genes have been linked to the occurrence of mutations that inactivate the *p53* gene, a tumour suppressor gene (Ryk *et al.*, 2006; Whibley *et al.*, 2009). Wu *et al.* (2006) found that women with a particular combination of alleles at two detoxifying genes, cytochrome P450 and glutathione S-transferase, had significantly increased risk of developing breast cancer, even without smoking or drinking habits. Furthermore, there is evidence that genotype-

specific susceptibility varies among different human ethnic groups (Bhisey *et al.*, 2003; Chikako *et al.*, 2005; Distelman-Menachem *et al.*, 2009).

There are differences in susceptibility (measured as EROD expression and liver tumorigenic damage) between flatfish species. For example, in the Puget Sound, USA, that was heavily polluted with PAHs (polycyclic aromatic hydrocarbons derived from the incomplete combustion of organic matter (Srogi, 2007)), both English sole, *Pleuronectes vetulus,* and starry flounder, *Platichthys stellatus*, are sympatric. The English sole has a high incidence of liver neoplasia (Johnson *et al.*, 1998; Myers *et al.*, 2008) and is routinely used as a bioindicator of pollution; whereas for the starry flounder very few instances of liver tumours have ever been recorded in the latter species (four in 20 years) (Reichert *et al.*, 1998). The different ways in which the species metabolise pollutants, combined with variance in cytochrome P-450IA (CYP1A) expression is thought to be responsible for the susceptibility to the development of liver pathologies (Reichert *et al.*, 1998). The CYP1A gene has a role in neutralizing toxic compounds (i.e. PAHs) (Carajaville *et al.*, 2003) and fish exhibit high CYP1A activity after exposure to pollutants (measured as EROD activity). However, elevated CYP1A activity is also associated with production of carcinogenic reactive oxygen species (ROS). Therefore there are both positive and negative effects of having a highly active CYP1A gene. Wirgin *et al.* (1991) revised the CYP1A genetic profile of PAH-exposed and unexposed populations of tomcod, *Microgadus tomcod*, and found evidence of polymorphism at the mitochondrial genes regulating CYP1A for the first time in teleost fishes, opening the way for studies on selection on CYP1A. Populations of killifish, *Fundulus heteroclitus,* chronically exposed to PAHs show evidence of a refractory CYP1A response (low EROD inducibility) (Meyer *et al.*, 2002). Low EROD response in killifish from polluted sites was inherited at least to the F1 generation, suggesting a genetic component in CYP1A activity; however, high EROD inducibility was recovered after two generations (F2 and F3) reared in uncontaminated environments (Meyer *et al.*, 2002), which contradicted the genetic hypothesis. Studies on European flounder, *Platichthys flesus*, using real time PCR and microsatellites, identified the quantitative and qualitative diversity of CYP1A-coding genes at the intra- and inter- family level (Dixon *et al.*, 2002), and David *et al.* (2003) found evidence that different populations of *Daphnia pulex* exposed to different levels of polyphenolic compounds in their environment, showed different profiles of another cytochrome P450, the CYP4 gene family. Overall these studies suggest that populations can change their profiles of biomarker response (i.e. EROD induction) after long term exposure to pollutants, thus indicating evolutionary adaptation of populations to pollution.

In the case of dab, *Limanda limanda*, some locations consistently show higher frequencies of liver pathologies than others (CEFAS, 2003a; 2005; Feist *et al.*, 2008; Stentiford *et al.*, 2009). Although in principle they are associated with pollutant exposure, the relationship is not always so simple. Fish collected from certain locations, such as Cardigan Bay, with no nearby sources of industrial and domestic pollutants (MAFF, 1990; CEFAS, 2000), suffer from high incidence of tumours in their liver (Lyons *et al.*, 2006; CEFAS, 2003a; Feist *et al.*, 2008). On the other hand, dab from sites such as Amble located only 30 km from the historically polluted Tyne estuary (Hudson-Edwards *et al.*, 1996), have rather low frequency of liver lesions (Feist & Stentiford, 2005; Feist *et al.*, 2008). There are many possible explanations for such patterns, such as age, environmental quality of nursery grounds, migration between locations, or genetic differences in individual response to pollutants (Kirby *et al.*, 2006). The latter factor could also be population-specific, and could be influenced, through selection, by previous exposure to pollutant. Changes in allele frequencies in fish populations have already been reported from polluted environments (Wirgin *et al.*, 1991; Theodorakis & Shugart, 1997; David *et al.*, 2003; Bourret *et al.*, 2008), and may have been caused by selection of particular individuals carrying specific alleles at certain genes. If populations can adapt to high pollution exposure, they might show reduced biomarker response (Meyer *et al.*, 2002; Kirby *et al.*, 2006). Therefore, it is important to evaluate whether dab has adapted to life in a polluted environment, as such process would impinge on the interpretation of biomarker response.

5.2.1 Genetic markers to detect selection:

The increased availability of Expressed Sequence Tags (ESTs), even for non-model species, has attracted the attention of molecular ecologists. Their assumed coding nature and thus potential for being under selection, together with the possibility of assigning protein function, offer an ideal counter-balance to the putative neutrality of microsatellite markers (Vasemägi *et al.*, 2005; Bouck & Vision, 2007; Naish & Hard, 2008). If a particular point mutation endows a net benefit to the carrier, then the offspring will not only inherit the point mutation, but large sections of DNA around it, depending on where the nearest recombination points reside. The process is known as genomic hitch-hiking (Smith & Haigh, 1974) and extends the signal of selection from the single point mutation to a more sizeable area of the genome (Figure 5.1). If selection is strong enough, the area may become a *variability valley* or *sweep detection window* (Teschke *et al.*, 2008), and markers within that area would show evidence of selective sweeps (Guinand *et al.*, 2004) (Figure 5.2). Two types of markers can be used to find such variability valleys: single nucleotide polymorphisms (SNPs) or microsatellites (single sequence repeats = SSRs). The former offer the advantages of being very common and

possibly identifying the selected site, but selective sweeps will be harder to detect due to the lower variability of the marker. Microsatellites will be present in lower frequencies throughout the genome, but their higher allele variability significantly increases their potential for detecting changes in allele distribution due to recent selective sweeps relative to a correspondingly similar number of SNP loci (Schlötterer, 2002; Teschke *et al.*, 2008).

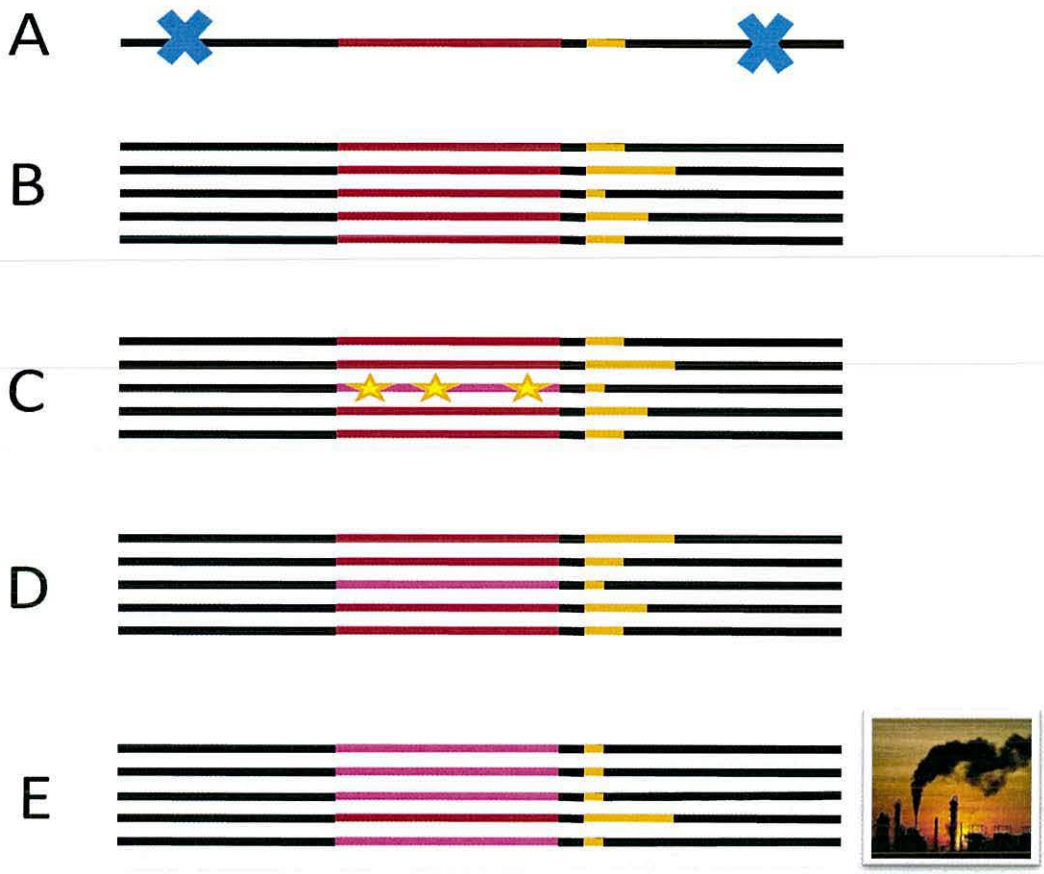


Figure 5.1: Diagram of a selective sweep and the principle behind EST-SSRs

The diagram represents a DNA sequence as line (**A**). Within the line there is a functional gene coloured in red. Closely linked to the functional gene there is a repetitive element (a SSR/microsatellite) coloured here in yellow. Both the functional gene and the SSR are inherited together as recombination (depicted by the blue crosses) occurs at either side of the gene-SSR duo. A population of individuals is represented (**B**). All individuals have the same red gene, but polymorphism in the number of repeats in the SSR (length of the yellow line) is present among the individuals. Mutation (yellow stars) might occur within the functional gene (**C**). Most times this will lead to a failure in functionally and the disappearance of the new mutation, but sometimes the new polymorphism might not imply negative effects, and a new allele emerges at the functional gene (coloured in pink). If the new "pink" allele does not entail any advantage in the current environment, its frequency within the population will be a product of drift and migration (**D**). However, under novel selective pressures (e.g. pollution), previous relatively neutral mutations (or even slightly deleterious) might become decisive in survival and rapidly "sweep" across the population and become the dominant allele (**E**). Note that the SSR-allele linked to the selected functional allele has now become nearly fixed.

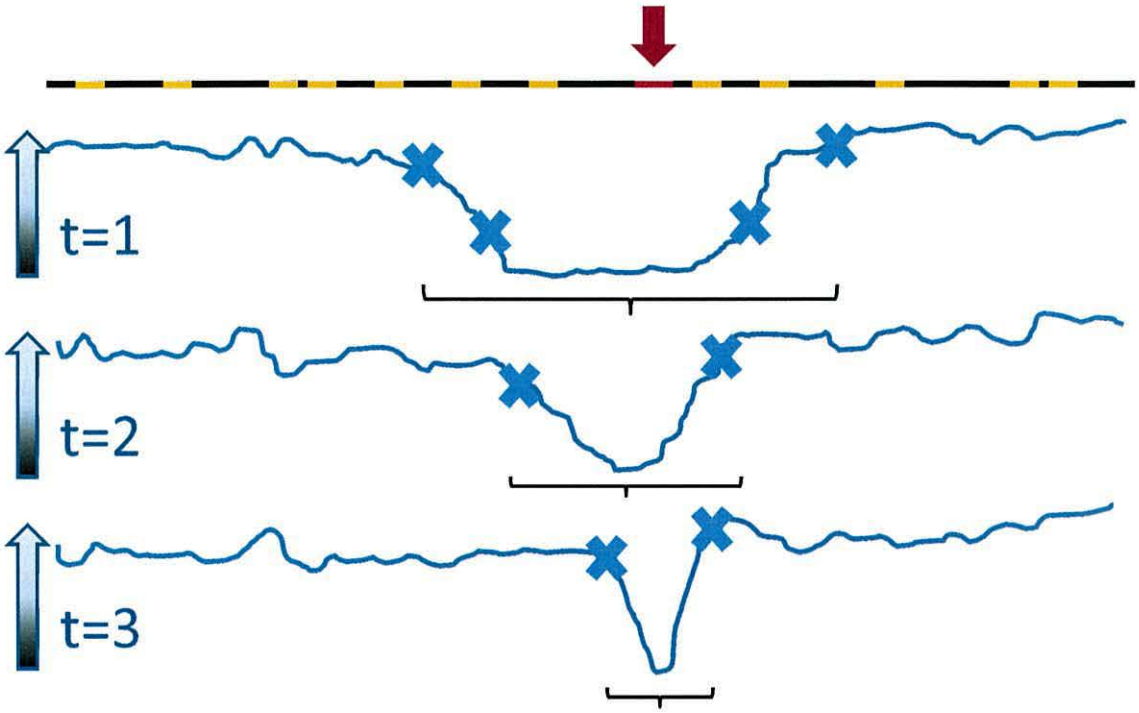


Figure 5.2: Diagram of reduction in variability around a selection point.
The black line represents a sequence of DNA in which a particular functional gene is under strong selection (gene in red and pointed to by a red arrow). Several genetic markers, either microsatellites or SNPs, are genotyped around the selection point. The level of polymorphism (blue arrows) of the genotyped markers is indicated by the blue trace: the higher it is the more polymorphism is detected at the marker. In the first trace, selection has just started operating and a large section of the DNA is inherited together with the functional gene, thus creating a *variability valley* and a *sweep detection window* (black bracket). However, in further generations (t=2, t=3) recombination events will start to break the linkage between the selected locus and the genotyped markers, reducing the width of the *variability valley* and the *sweep detection window*.

Microsatellites may be located in coding regions where their polymorphisms affect protein function or even cause disease (Cummings & Zoghbi, 2000); may not be located within open reading frames but may have regulatory functions over gene expression or be structurally important in DNA folding (Li *et al.*, 2002); or may have unknown function within the EST (Li *et al.*, 2004). Microsatellites derived from EST libraries are known as EST-SSRs (Expressed sequence tags simple sequence repeats) (Ellis & Burke, 2007), and conventionally have been genotyped in the same way as putatively neutral microsatellites and the resulting sample pair-wise F_{ST}'s screened for outlier loci (Beaumont & Nichols, 1996; Beaumont & Balding, 2004; Antao *et al.*, 2008).

The detection of outlier loci has resulted in two problems: in order to identify statistically significant outlier loci, a large number of loci need to be screened in many individuals. Only large changes in allele frequency (a population is monomorphic while others are not) will be

identified as significant (Vasemägi *et al.*, 2005). Secondly once an EST-SSR is found to be under selection, it is usually not annotated and no function can be allocated to it, thus the nature of the selective pressure remains unknown (Vasemägi & Primmer, 2005; G. Hoarau, personal communication referring to Coyer *et al.*, 2009). In order to address the first problem, DNA pooling (Breen *et al.*, 1999; Ritland, 2002) has been suggested as a quick way of evaluating selective sweeps (Thomas *et al.*, 2007). Essentially, individuals are pooled according to provenance or any other characteristic, and screened for a multitude of markers, in a search for selective sweeps or allele range changes. The problem on non-annotation can be overcome if EST-SSRs are restricted to those for which a function has been allocated.

Several studies have examined differences in expression profiles of flounder before and after exposure to pollutants, herbicides and other environmentally relevant chemicals (Sheader *et al.*, 2004; Sheader *et al.*, 2006; Marchand *et al.*, 2006; Williams *et al.*, 2007). These studies have found many genes that were up or down-regulated after treatment compared to controls, suggesting their involvement in pollutant response. Of the nearly 8000 ESTs published, a few (ca. 600) were identified by homology to actual proteins and known functions, and were classified into three broad areas: Energy-related, detoxification, and tissue injury/tumour growth (Marchand *et al.*, 2006). Here, microsatellite-embedded ESTs are used to test whether a relationship exists between polymorphism in differentially expressed genes, considered to be relevant for the individual success in a polluted environment, and biomarkers of pollution exposure, interpreted as a proxy of success/failure in a polluted environment.

5.3 Materials and Methods:

5.3.1 EST-SSR development:

Here, only the differentially expressed ESTs after exposure to pollutants (Sheader *et al.*, 2004; Sheader *et al.*, 2006; Marchand *et al.*, 2006; Williams *et al.*, 2007) were used in order to maximise the chances of finding pollutant-relevant genes. The published flounder EST library was screened for microsatellite motifs using the software SPUTNIK (Abajian, 1994). Roughly 500 sequences contained microsatellite-like motifs. The resulting microsatellite library was cross-searched for known-function ESTs, and the ensuing ESTs blasted against the NCBI database (http://www.ncbi.nlm.nih.gov) to compare the microsatellite region of the sequence with other homologous sequences. If the microsatellite region was monomorphic throughout a wide range of taxa, the EST was rejected. The selection process reduced the numbers to 31 known-function potentially polymorphic EST-SSRs. A further four unknown

function EST-SSRs were selected, based on their high probability of polymorphism. Primers were designed and tested as in Chapter 2 and 3. Primers were tested for amplification on individual flounder and individual dab. Amplification success, first assessed on 3% TBE agarose gels, was high: 30 EST-SSRs amplified at the expected size range in flounder and 31 amplified in dab (Table 5.1)

5.3.2 Experimental design:

A sample of dab was collected from the North Sea as part of the CSEMP biomonitoring programme in North East Dogger (NeD07), a location that has a very high prevalence of liver lesions (Stentiford *et al.*, 2009). Another sample of fish collected in Irish Sea in 2007 with low prevalence of liver lesions was used as a comparison.

Two contrasting liver lesions were chosen: Hepatocellular adenomas and lipoidosis. The former are benign neoplasms and were chosen because of their potential to develop into life-threatening condition (carcinoma), relevance to human health studies, and their known causality by pollutant exposure (Baumann, 1998; Feist *et al.*, 2004; Koehler, 2004). The latter is the abnormal retention of lipids on the liver which can be chemically induced, and is thought to be an early toxicological response (Köhler *et al.*, 1992), is commonly found in fish from heavily polluted areas (Teh *et al.*, 1997; Schlacher *et al.*, 2007), and can be fatal (Penrith *et al.*, 1993). In the current study, the aim is not to identify genes *directly* linked to the propensity of developing a disease (i.e. whether the disease is heritable), but to test whether certain alleles among the EST-SSRs considered are associated with alleles at genes that either improve or hinder detoxification mechanisms (i.e. variability in the differentially expressed genes leads to variability in biomarker response).

Table 5.1: EST-SSR testing **(Next page)**

Testing EST-SSR for evidence of selective sweeps in dab, *Limanda limanda*. **EST** = GenBank accession number; **Motif** = microsatellite motif found in sequence; **Function** = putative function identified by homology; **F-primer** = forward primer for microsatellite amplification; **R-primer**= reverse primer for microsatellite amplification; **Size** = expected product size; **EST-SSR** = Name of the gene-associated microsatellite; **F** & **D** = test of amplification in flounder (F) and dab (D) on agarose gels: 1= success, ~= unresolved, - = negative. **Poly?**= test for allelic polymorphism, mono= monomorphic, yes = polymorphism was found, ? = unresolved, no amp = no amplification; **size range** = allele size range found for the locus; **µsat?**= whether the microsatllite was deemed usable as a population marker; **Differences?** = test of whether there were differences in the pooled amplifications between the different phenotypes considered: (**A**) fish from Dogger Bank suffering from liver adenomas; (**B**) fish from Dogger Bank suffering from lipoidosis; (**C**) fish from Dogger Bank with no recorded liver diseases; (**D**) fish from the Irish Sea with no recorded disease.

EST	Motif	Function	F-primer	R-primer	Size	EST-SSR	F	D	Poly?	size range	µsat?	Differences?
DV565603	(ATT)15	Additional Sex Comb-like 1	TTGGTCAGAGCACACAGGAG	GCCACCAAAAGATGACAGGT	250	EST-01-N	1	1	mono	270		no
AJ578051	(AC)21	apolipoprotein A1 precursor	ATCCTCTCCACCCTGTTCCT	CTGTTCTCCAAATTCTCCTTAGAG	162	EST-02-N	1	1	yes	130-167	?	Reduced var C&D
EC378204	(AGA)8	Binding prot/Peptidyl-proline isomerase	CCGATCCATTTTGGACAGAC	TGAGCAACAACTTCCGTGTC	119	EST-03-N	1	1	mono	111		
EC378419	(GT)8	Cell Death Apoptosis 10	CCAGCCCGATTAAGAAGTCA	GGAGCAGCTCTTGGTCAAAC	154	EST-04-N	-	-	4peaks	149-157	?	no
DV568872	(AGG)6	Cell Death Apoptosis 4	AACAGCTGAGGGATGCTTGT	GGCAGAATCAAGGCGTTAAG	156	EST-05-N	1		?		?	?
DV569360	(AC)16	claudin	GGGGGTCATGTTCATTCTGT	GTGAGCGTGAAGACACCTGA	185	EST-06-N	1	1	yes	167-182	yes	no
EC378628	(CT)11	COMM 1/copper metabolism	GCCAAAGTGAACCAGATGCT	GAGTCCGCTCTGGTCTGAAC	175	EST-07-N	1	1	mono	191		
EC379377	(AC)16	Complement control protein	TCTCTTACCTGTGGGGCATC	CAACTCAAGTTAGAAAAGCAACG	216	EST-08-N	1	1	yes	292-302	?	no
EC378550	(AC)10	Complement control protein factor	TCAGAAGCAACACAGGTTGG	TTGACAATTTTGGGAATGTGA	158	EST-09-V	1	1	yes	217-232	yes	?
CF379162	(AC)9	Cyt malate dehydrogenase	CTCCTGGACGGCTAATGTGT	CCCTGAAGACTTGCTGTTCC	234	EST-10-V	1	1	yes	232-246	yes	no
DV568350	(AT)6	Cyt P450 CyP2F2	CGGGCAGAATTTTGAATGTT	CAAATACAGGCAGCAAGCAA	196	EST-11-V	1	1	yes	175-185	yes	A has a longer allele, 185
EC377805	(GT)8	cytochrome c oxidase copper chaperone	AGAGGAAAGCTGCACAGAGC	TAAATAAGCGGGGGAAAAGG	152	EST-12-V	-	-	mono	133/140/1	?	no
EC378538	(TG)8	Flavin adenine dinucleotide synthetase	CGCCTCACACCTCATTACCT	GCCATTACTTCAGACGCACA	195	EST-13-V	1	1	mono	212		no
DV568641	(CCT)8(GT)8(GT)14	GTP Cyclohydrolase 1	CGAGTTTCTGACGCTGATCC	TGTCCAGAATCGTTTGACCA	219	EST-14-V	1	1	yes	230-234	yes	no
DV565618	(GGT)7	Hsp40	GGGGAAGAAGGTCTGAAAGG	CCCCGAAGAACTGTTCAAAG	151	EST-15-V	1	1	mono	159		no
CF379117	(TG)6	Inmunoglobulin M	CTCAACCTGGTCAACGTCAG	TTTTATTTTGATTTTGAATCTGCAT	152	EST-16-V	1	1	mono	166		no
EC379462	(AC)22(CT)5	MHC class 1 Antigen	GCCCTCAAACAAATTTCCTC	GGGGCTGTTCTCCAAATTCT	172	EST-17-V	~	1	yes	213-238	yes	no
CF379204	(AG)x (CT)x	Myosin Light Chain	GCCGGCAACGTAGACTACAA	TGAGTGGCTGAGCACATAGG	187	EST-18-P	1	1	mono			no
DV569527	(AAG)43	p8 protein (metastasis)	CAGATCCTGTTGTTGTTCTGCT	CCACGTGACTCATGTTGGTC	209	EST-19-P	1	1	yes		yes	Reduced var B
EC379653	(GT)6	Pleurocidin (antimicrobial)	AGCTAGAGCAGGGGTTTTCA	CCAACATGGAAAACCAAATGT	166	EST-20-P	1	1	?		no	no
EC378887	(GT)18	Profilin	CCACCAATCCTGATGGCTAC	ACAGGCAGCTCAGGTGTGTA	212	EST-21-P	~	1	?		no	no
DV568031	(AC)5	Protein Lysine	TTGCTTCTCGTATCCCCTCT	AAGTGGTTCATTCAAAAAGTGTGA	104	EST-22-P	1	1	mono	122	no	no
DV565797	(ATC)9	Selenoprotein P	GACATCTGCAACTGCTCTGC	CCCCTGAATGTTATGATGGTG	207	EST-23-P	1	1	mono	198	no	no
DV567967	(TG)10	Sulfated glycoprotein (tumour)	GCCTTTGAGCACATTTGGTT	GCGTCGCCAACAACTTTAAT	108	EST-24-P	-	1	yes	116-144	yes	A only 144, others 124-144
DV570271	(AG)X	Syndecan 2	GAAACAAAAACAAAGAGCAGGA	GACAGGCGTCAAAATGTCAA	232	EST-25-P	-	1	mono	240	no	no
EC378618	(GCT)7	Toxin 1	CGAACAGAAAACCAGAACCA	CTCACCGTTATCTCGGGTGT	178	EST-26-P	~	~	yes	203-221	yes	no
DV566291	(AC)20	Trafficking Prot subunit 1	ACGGCTTGGACAGTGAACTC	CACAACAGTTTGATTCATGTTGC	174	EST-27-F	1	1	yes	236-260	yes	Reduced var A&B?
EC378076	(GT)30	Trancobalamin I Precursor	GAGGGGGACAAGATCACTCA	GCATGATGCACACACTCACTC	235	EST-28-F	1	1	yes	245-291	yes	?
CF379224	(TGC)4	translation elongation	ACAAGGTTGGAACCGACATC	GCTCAAAGCTTTAATAACCGTGT	171	EST-29-F	~	~	no amp			
DV569969	(GT)19	Ubiquitin conjugating Protein	CCACTCCCCTGTTAAGTTGC	TAAGGCAAACGGTGGAAATC	194	EST-30-F	1	1	yes	185-205	?	Reduced var B
DV567625	(CAG)5	Vitellogenin A	ATCATCCTCCTCCAGCTCCT	GAGAGCGTTGACCCTGTTTC	188	EST-31-F	-	-	mono	189-211	no	no
CF379086	(GT)9	z-unknown function	GATCCAGGCATCTGTGGTTT	AAACAGGCAATACAAAATGAGAAA	120	EST-32-F	1	1	mono	129	no	no
CF379071	(AG)10	z-unknown function	ATGTGTCATCGCTGTCTTGC	CTGGTCTGACAGTCGTTCCTT	151	EST-33-F	1	1	yes	173-183	yes	no
EC379649	(AG)23	z-unknown function	ACGAGAGGCCCCTTTAAATC	TCGAACGACGAATGTATGGA	240	EST-34-F	~	~	mono			no
EC379543	(CT)6(AC)21	z-unknown function	TGGACTCCCAGTCTCCTCTG	CAGCAGCAGCAGTCAAGTTT	155	EST-35-F	1	1	yes	185-191	yes	no

Four different "Phenotypes" were considered:

A) Fish with hepatocellular adenomas from NeD07: "Adenoma NeD fish"

B) Fish with lipoidosis but no tumorigenic activity from NeD07; "Lipoidosis NeD fish"

C) Fish with no recorded liver disease from NeD07; "Healthy NeD"

D) Fish with no recorded liver disease from Liverpool Bay (LiV07): "Healthy LiV" Bau98

Evidence for selective sweeps at any of the EST-SSRs could occur in three different ways: First, a reduction of allelic diversity in either of the disease phenotypes would imply that certain alleles are less effective at detoxification, rendering the individual susceptible to disease; Second, allele fixation in fish with no disease record from NeD07 would suggest that the allele is linked to a key gene that is contributing to disease resistance in a noxious environment. And finally, a selective sweep could have affected a whole NeD07 population, hence the inclusion of fish from the Irish Sea, which was used as a control to confirm that any apparent lack of polymorphism at in NeD07 phenotypes was not due to a pan-population selective sweep. DNA from 22 individuals from each of the "phenotype" groups were normalised to 50ng/µl and pooled together. Additionally four individuals were also scored at each EST-SSR to assess peak shape and polymorphism. Standard PCRs were performed as previously described in Chapter 4 with tailed-forward primers and labelled tail-oligos. Different EST-SSRs were not multiplexed or poolplexed initially to avoid confusion with cross-colour pull-ups (an artefact peak created by an interference between spectral absorbances) (Rudin & Inman, 2002). The genotypes were analysed using GeneMapper®.

5.4 Results:

Most ESTs isolated from flounder, amplified in dab. Of those, 13 were monomorphic and one had three peaks per individual, but showed no variability between individuals. The remaining 15 showed polymorphism among individuals and showed several peaks in at least one of the pooled amplifications.

Seven of these EST-SSRs showed differences among pooled DNA phenotype groups:

EST-02 (apolipoprotein A1 precursor): On the first amplification, Adenoma and Lipoidosis fish showed three peaks, while healthy fish from NeD and LiV only showed one peak, although with different sizes in base-pairs (bp) in each case. Weak amplification in single individuals suggested the presence of null alleles. The forward primer was redesigned (but there was no space on the flanking region for redesigning the reverse primer). The new primer combination improved microsatellite patterns, albeit null alleles were still prevalent. No differences among phenotypes were found between groups with the new primer combination.

EST-11 (COMM 1/copper metabolism): On the first amplification, three alleles (176, 182 and 185pb) were identified for Adenoma fish, but only two (176 and 182bp) for the rest of phenotypes.

EST-14 (GTP Cyclohydrolase 1): Healthy fish from NeD showed reduced variability with only two alleles (232 and 234bp), while Adenoma fish showed a large range of alleles from 206 to 242bp, lipoidosis and healthy fish from LiV showed intermediate ranges of alleles (212-234bp).

EST-19 (p8/Metastasis): Failed to amplify in the lipoidosis fish, while other amplifications were successful. The PCR was repeated to confirm the pattern. Single amplifications revealed that most individuals from the Lipoidosis group failed to amplify, but failed amplifications in the other groups were also found. Mispriming was suspected as the source of the pattern, thus a new reverse primer was designed and tested (there was no space in the flanking region for redesigning a forwards primer). The new primers amplified successfully in all samples, thus confirming problems with the first reverse primer and rejecting the possibility of a locus deletion.

EST-24 (Sulfated glycoprotein/tumour inducer): Initially, only one allele (144bp) was found in the adenoma fish, while the other groups revealed several alleles. Eight individuals per phenotype were then genotyped singly, which revealed that individuals showed up to four peaks suggesting the primers were not only amplifying the target gene. PCR conditions were made more stringent by increasing the annealing temperature to 58°C and reducing the number of cycles to 25. Four individuals with adenomas and four healthy individuals from NeD were genotyped. With the new PCR conditions, the patterns of variability were reversed: individuals with adenomas showed two alleles (142 and 144), while healthy individuals showed only one allele (142).

EST-27 (Trafficking Prot subunit 1): fish suffering from adenomas and lipoidosis showed a skewed distribution of alleles towards shorter alleles compared to all other healthy fish. However, single amplifications were distorted by a non-specific band so new primers were re-designed and tested. The new primers showed no difference in allele distribution.

EST-30 (Ubiquitin conjugating Protein): Three peaks (185, 192, and 205) were present in fish with adenomas and healthy fish from both NeD and LiV, but peak 192 was missing from fish suffering lipoidosis.

Overall, null alleles were detected at several loci which resulted in a few initial false positives. After primers were redesigned for the null allele-affected loci, no marker showed signals of recent strong selective sweeps (i.e. lack of polymorphism) in any of the phenotype groups.

However, a few loci showed reduced variability in at least one of the groups compared to the others.

5.5 Discussion:

The development of microsatellite markers from EST libraries proved successful. Even when the ESTs had originally been isolated from another closely related species, most amplified in dab, suggesting that targeted regions of the dab and flounder genomes are similar. Twelve of these EST-SSRs showed good microsatellite profiles and, being polymorphic, would probably be well suited as population markers. Microsatellites developed from EST libraries are now commonplace in population genetic analyses (Rossetto *et al.*, 2002; Casas *et al.*, 2005; Stenvik *et al.*, 2006; Ellis & Burke, 2007; Coyer *et al.*, 2009), thus the EST-SSRs developed here should also be considered in future dab (or flounder) genetic studies. Although the pooling approach proved a good way of assessing patterns of polymorphism, it does not negate the need for individual screening of large numbers of individuals for many loci in order to detect weaker or older signals (Teschke *et al.*, 2008). Two key problems to the approach emerged in the current study:

First, the pooling approach is valid when screening large numbers of loci (ca. 1000 mouse microsatellites for the Thomas *et al.* (2007) study), where the focus is on detecting very strong and recent selective sweeps which lead to the fixation of only one allele. However, the *extended* signal of selection across variability valleys becomes eroded across generations (Figure 5.2), as recombination reduces the extent of the "valley" and variability is incorporated around the selection point, even if selection is still active. The rate of polymorphism recovery will be locus-specific depending on its unique mutation characteristics (Brandström & Ellegren, 2009). As a result, the power of the approach decreases over time since the selection event. The power also decreases when trying to detect weaker selective pressures: For example, when the selected allele does not become fixed for the population (Smith & Haigh, 1974), or where the target gene is quantitative in nature (Glazier *et al.*, 2002).

Secondly, null alleles (due to either primers failing to anneal or preferential amplification of certain alleles in the pooled reaction) were present in a few loci resulting in false positives. The mispriming seemed more prominent in the pooled samples that in single individuals, in accordance to results reported by Teschke *et al.* (2008).

In the current study, some markers showed reduced variability, but not complete allele fixation, in one or more of the phenotypes compared to others (EST-11, EST-14, and EST-30). Such observations suggest the possibility of selection for or against specific alleles in nearby loci. Nevertheless, given there is still variability in these loci, larger numbers of individuals would need to be singly genotyped to evaluate unambiguously a selective response at the locus and gain statistical significance. Such an individual genotyping could be done in conjunction with neutral microsatellites.

To conclude, the limitations imposed in the current approach, i.e., only using ESTs involved in the toxicological response of fish, may overly reduce the size of the initial test panel. In particular, using only loci for which a function could be allocated reduced our study to using genes within the flatfish genome that are conserved across taxa, while some of the most interesting adaptations to pollution tolerance might be unique pathways evolved within flatfish, and thus unreported from other taxa. Furthermore, individual variability in gene expression may be less related to mutations in the measured gene, but in gene promoters in completely different parts of the genome (*trans*-acting factors) (Vasemägi & Primmer, 2005). Finally it should be noted that the lack of evidence of selection, is not evidence that it does not exist, but that more comprehensive approaches to evaluate pollution selection pressures are required.

5.6 Summary:

- Thirty-five microsatellite markers linked to genes differentially expressed after exposure to pollution were designed and tested on groups of healthy and diseased fish.
- Of the polymorphic markers, none showed complete fixation for a particular allele in any of the fish groups in the pooled amplifications. However, reduced polymorphism was found in some of the loci-group combinations.
- When using the EST-SSR approach, the exclusion of un-annotated sequences is not recommended, as it limits the sampled genome to phylogenetically conserved areas while recent and more relevant adaptations might be unique to the studied taxon.
- Individual genotyping of fish with the polymorphic markers, in combination with neutral markers, is suggested as a means to evaluate the possibility of weak/older selection events at these loci.

Chapter 6: Differentiation in genetically diverse marine populations: an empirical comparison of heterozygosity-tolerant estimators

6.1 Abstract:

Estimating the partitioning of genetic variability among wild populations continues to play a pivotal role not only in evolutionary biology and ecology, but also in conservation biology. Accordingly, it is important to employ an appropriate estimator of genetic differentiation that matches the salient attributes of a genetic marker system. Only when certain conditions are met can differentiation estimators generate meaningful interpretation of patterns of population connectivity and migration. Although there is an extensive array of estimators available based on various models of population structure and mutational mechanisms, the effect of heterozygosity on such estimators have, until recently, received little attention. The key issue concerns the limit imposed by heterozygosity on the maximum value of differentiation between populations that can be estimated. Microsatellites can have high allelic diversities with correspondingly high heterozygosities, and in marine organisms, which typically comprise large effective population sizes, heterozygosities can reach problematically high levels. To address issues arising from high heterozygosities, a new generation of estimators has been devised: G'_{ST}, φ'_{ST}, Θ'_{WC}, and D_{est}. In the present study, the performance of traditional (Θ_{WC}, G_{ST}, φ_{ST}) and "high-heterozygosity-tolerant" estimators are compared by application to a large spatially and temporally replicated microsatellite data set from dab (*Limanda limanda* L.): a European flatfish with large population sizes and recent colonization history - features typically resulting in high genetic variability and subtle genetic structuring. There was concordance in the ranking order of information content of loci with all heterozygosity-corrected estimators, though with marked differences to traditional estimators. Contrary to common perception, highly diverse microsatellites proved more informative in elucidating population genetic structure than reduced variability ones. Such discrepancies emphasize the inherent limitations of traditional estimators in understanding the distribution of information revealed by highly diverse markers. In conclusion, the use of "high-heterozygosity-tolerant" estimators is recommended for biological systems where markers are characterised by high and variable levels of heterozygosity across loci.

6.1.1 Glossary of terms and abbreviations:

Chapter 6 deals with a set of complex statistical concepts relevant to the analysis of population genetic data. To simplify the reading, the definition of some commonly used terms and the explanation of abbreviations used are provided below.

Heterozygote: an individual with two different alleles for the same locus

Homozygote: an individual with two equal alleles for the same locus.

HWE: Hardy-Weinberg Equilibrium. The Hardy-Weinberg model states that, if two alleles of the same locus, A and B, are present in frequencies p and q (where $p + q = 1$), then the frequencies of homozygotes for A and B are AA = p^2 and BB = q^2, while heterozygotes will be present at AB = $2pq$; (note that: $p^2 + q^2 + 2pq = 1$). In the absence of selection, non-random mating and gene flow, the genotype frequencies remain essentially unchanged across generations. The model can be extended for loci with more than two alleles following the same principle, and larger number of alleles in a locus results in higher expected frequency of heterozygotes under the HWE.

H_O: Observed frequency of heterozygotes at a given locus.

H_E: Expected frequency of heterozygotes at a locus under HWE given the allelic diversity at the locus. Due to its tight relationship with number of alleles it is also known as gene diversity.

H_S: Within-sample Heterozygosity: Expected heterozygosity (H_E) within a single sample.

H_T: Total heterozygosity: Expected heterozygosity (H_E) calculated from a combination of samples.

UCE: **Uncorrected** Estimators of genetic differentiation: These are estimators of differentiation used in their native state which do not include a correction for high heterozygosity. The category includes Θ_{WC}, G_{ST_est}, and φ_{ST}.

HCE: Heterozygosity-**Corrected** Estimator of genetic differentiation: these are estimators derived from UCEs for which a correction for heterozygosity has been applied. The category includes Θ'_{WC}, G_{ST_est}', and φ'_{ST}. Note that HCE are always indicated by a (').

HTE: Heterozygosity-**Tolerant** Estimator of genetic differentiation: these are both estimators corrected for heterozygosity (HCE) and estimators purposely constructed to take into account differences in locus heterozygosity (i.e. D_{est})

Δ_{HC}: Locus-specific magnitude of change between UCE and HCE estimates, which was calculated for each locus, year and pair of Θ_{WC} and Θ'_{WC} as: $\Delta_{HC} = \Theta'_{WC} - \Theta_{WC}$.

Homoplasy: within a locus, the mutation of an allele to another already pre-existing allelic state, thus not resulting in a new detectable allele. The occurrence of homoplasy is problematic for population genetic studies as the resulting two alleles are not distinguishable by detection methods (identical by state) but are not related to each other (identical by descent).

Null allele: an undetected allele, either due to mutations in the priming site or scoring errors. Null alleles may increase the frequency of apparent homozygote individuals and disrupt the HWE.

6.2 Introduction:

The quantification of genetic differentiation among populations has generated much discussion ever since Wright devised measures to describe the distribution of genetic variation in wild populations (1943; 1951). The level of structure between populations is defined by a parameter (i.e. F_{ST}, D, R_{ST}) which is the *real* value of structure associated with the studied organisms and that, as all parameters, can never be truly known or computed, even if the whole species is genotyped. Instead, geneticists use differences in genetic diversity among *samples* collected from the populations thought to be representative of those populations; statistics are then used to reach an approximate value of the *real* parameter (Weir & Cockerham, 1984; Hedrick, 1999a). Statistical values, when used to evaluate a parameter, are known as estimators.

Although there are a plethora of genetic differentiation estimators, a few have dominated the field of population genetics: D_S and unbiased D_{SU} (Nei, 1972; 1978), D_C (Cavalli-Sforza & Edwards, 1967), G_{ST} (Nei, 1973), Θ_{WC} (Weir & Cockerham, 1984) and φ_{ST} (Excoffier *et al.*, 1992) are most commonly reported. Many of these estimators were initially developed to analyse trait frequencies, blood groups, allozymes and sequence data which are all characterised by low to moderate levels of diversity. However, the discovery of highly polymorphic microsatellite markers has revolutionized the field of population genetics, increasing the power and scope of questions that can be addressed (Jarne & Lagoda, 1996; Li *et al.*, 2002), especially in weakly differentiated systems such as marine fish (Carvalho & Hauser, 1994; 1998; Chistiakov *et al.*, 2006; Hauser & Carvalho, 2008). As microsatellites became more commonly used, their high mutation rates and potential limit in number of alleles (Goldstein *et al.*, 1995a; Goldstein & Pollock, 1997) led to the development of several new estimators of genetic differentiation that take into account the mutation mechanisms associated with microsatellites, that is, the stepwise mutation model (SMM) and the limited K-allele model (KAM) (Kimura & Ohta, 1978; Shriver *et al.*, 1993; Valdes *et al.*, 1993). The new estimators were $\hat{R}_{ST}$ (Slatkin, 1995), $(\delta\mu)^2$ (Goldstein *et al.*, 1995b), D_{SW} (Shriver *et al.*, 1995) and $\widehat{\varphi_{ST}}$ (Michalakis & Excoffier, 1996). The performance of such estimators has been extensively tested (Takezaki & Nei, 1996; Goldstein & Pollock, 1997; Pérez-Lezaun *et al.*, 1997; Lugon-Moulin *et al.*, 1999), and although microsatellites do not conform to all the assumptions of the infinite allele model (IAM), estimators of F_{ST} (i.e. Θ_{WC}) generally provide a better representation of genetic differentiation within species than R_{ST} (Balloux & Lugon-Moulin, 2002; Kalinowski, 2002a; Estoup *et al.*, 2002a).

Most estimators of genetic differentiation are strongly dependent on the heterozygosity of the markers used. G_{ST} and to a lesser extent Θ_{WC}, were developed to detect differences in expected heterozygosities, and explain observed excess homozygosity by partitioning subpopulations (Nei, 1973). Therefore, differentiation cannot exceed the level of homozygosity (Hedrick, 1999b; Kalinowski, 2002a). With low heterozygosity markers such as allozymes there is considerable scope to assign excess homozygosity to substructure if needed (e.g. when samples do not comprise a panmictic population). However, as heterozygosity increases there is less scope (homozygosity) for partitioning, which results in negligible or no differentiation between samples (Hedrick, 1999b; Kalinowski, 2002b). Such patterns have been interpreted as highly heterozygous loci losing information content (becoming saturated) due to their high mutation rates resulting in homoplasy (Balloux *et al.*, 2000; Olsen *et al.*, 2004; O'Reilly *et al.*, 2004; Astanei *et al.*, 2005). It is generally believed that the potential for information content of a marker increases with heterozygosity until a maximum value is reached around H_E~0.700 after which the utility of a marker decreases sharply (Leblois *et al.*, 2003) to the point where Rousset (2008) recommends using loci with $H_E < 0.800$ for the computation of Θ_{WC}. Indeed, Olsen *et al.* (2004) suggested separating loci by heterozygosity and performing analyses separately so the effect of mutational bias in highly heterozygous loci could be ascertained. Taking account of these constraints on measures of genetic differentiation is of utmost importance for studies on taxa displaying high heterozygosities (Ward *et al.*, 1994; DeWoody & Avise, 2000). Indeed, the inverse relationship between Θ_{WC} and heterozygosity has been reported in several empirical (Paetkau *et al.*, 1997; Olsen *et al.*, 2004; O'Reilly *et al.*, 2004; Astanei *et al.*, 2005; Carreras-Carbonell *et al.*, 2006; Weetman *et al.*, 2006) and simulated studies (Kalinowski, 2002b; 2002a).

Populations with no shared alleles (i.e. indicating no recent genetic exchange), but high heterozygosity, may show little differentiation (Hedrick, 1999b; Carreras-Carbonell *et al.*, 2006). Therefore, highly heterozygous loci may not necessarily be saturated; instead, the estimators may be limited in their statistical properties at high heterozygosities. Focussing on these limitations on G_{ST}, Hedrick (2005) proposed a standardised G'_{ST} by dividing the estimate (G_{ST}) by the maximum hypothetical value that could be reached given the observed levels of genetic variation ($G_{ST(max)}$). Following the same approach, Meirmans (2006) modified an analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992) to obtain a similarly standardised differentiation estimator: φ'_{ST}. The latest addition to this family of HTE of genetic differentiation were devised by Jost (2008), where a new measure, D, and its unbiased estimator D_{est}, are suggested as measures of actual differentiation. Despite the

numerous studies on the effect of gene diversity on estimators of differentiation such as Θ_{WC}, relatively few studies have employed heterozygosity-corrected estimators of differentiation until recently (Heller & Siegismund, 2009). The initial uses of G'_{ST} were limited to the evaluation of outlier loci (Nielsen *et al.*, 2006), the comparison of different markers (Hemmer-Hansen *et al.*, 2007a), and testing especially whether maximum differentiation has been attained (Duftner *et al.*, 2006). However, the number of studies reporting standardised estimators of differentiation is increasing, targeting population structure of turbot, *Psetta maxima*, (Florin & Höglund, 2007), Trinidadian guppies, *Poecilia reticulata*, (Suk & Neff, 2009a), Antarctic icefish, *Chaenocephalus aceratus,* (Papetti *et al.*, 2009), several species of Tanganyikan cichlids (Wagner & McCune, 2009), American lobsters, *Homarus americanus*, (Kenchington *et al.*, 2009) and sea stars, *Astropecten aranciacus* (Zulliger *et al.*, 2009).

Although the new estimators of genetic differentiation are unlikely to radically change the patterns of significance of differentiation obtained with more traditional estimators, they may improve our understanding of *how* different populations are, and consequently improve our ability to identify mechanisms generating temporal and spatial patterns. In the field of fisheries genetics, for example, the sensitivity of Θ_{WC} to heterozygosity may explain, in part, the incidence of very low (Θ_{WC}= 0.002) but highly significant structuring commonly found in many marine fish (Lundy *et al.*, 1999; Ruzzante *et al.*, 1999; Knutsen *et al.*, 2003; Jørgensen *et al.*, 2005; Mariani *et al.*, 2005; Larsson *et al.*, 2007). Although there has been some discussion about the performance of the HTE retrospectively (Heller & Siegismund, 2009) and in theory (Ryman & Leimar, 2008; 2009; Jost, 2009), there is a need for empirical evaluation of the use of HTE and their implications in population structure interpretation and choice of markers. The aim here is not to revise the mathematical equations and assumptions underlying the estimators, which has been dealt with elsewhere (Hedrick, 1999b; Hedrick, 2005; Meirmans, 2006; Jost, 2008; Ryman & Leimar, 2008; 2009; Jost, 2009), but rather assess and compare the performance of the UCEs and HTEs of genetic differentiation on a temporally replicated data set from a weakly genetically differentiated marine fish species (large population sizes with recent divergence times and possible migration). The temporal replication facilitates assessment of the biological importance of any significant structure detected, while the inclusion of several spatial samples per putative population (basin) increases the robustness of the estimated differentiation values. Data are based on 16 loci showing a range of heterozygosity values, allowing evaluation of the performance of estimators at different levels of allele diversity.

6.3 Materials and Methods:

6.3.1 Sampling:

Genetic samples of dab were collected for four consecutive years (2005-2008) as part of the Clean Seas Environment Monitoring Programme (CSEMP) (CEFAS, 2005) in up to 15 stations covering five areas in the North East Atlantic. Although in later chapters these areas will be considered separately, in this chapter, for simplicity, two broad groups will be used and referred to as North Sea (North Sea and English Channel) and Irish Sea (Irish Sea and Atlantic). In total, 40 samples, each with 21 to 183 individual dab, *Limanda limanda*, were analysed (Table 6.1). DNA was extracted from fin clips using the hi-salt extraction method (Aljanabi & Martinez, 1997), and samples genotyped for 16 microsatellite loci published in Tysklind *et al.* (2009b) in multiplex PCRs as indicated in Chapter 4. Genotypes were determined with GENE MAPPER® and each individual genotype checked at least twice, and any rare alleles or HWE and LD outliers were checked for allele miscoring, size-standard miscalling, or cross-colour pull-ups (an artefact peak created by an interference between spectral absorbances, Rudin & Inman, 2002). In one sample, ScB06, DNA was extracted and genotyped twice, at the beginning and end of the project to check for scoring consistency (DeWoody *et al.*, 2006). At least two individuals per plate were re-genotyped to verify amplification consistency between plates. Some individuals could not be fully genotyped even after several attempted amplifications and have been removed for some of the analyses.

6.3.2 Data Analysis:

6.3.2.1 Power, locus characteristics and HWE:

The power of the microsatellite suite to detect differentiation between two hypothetical populations was assessed with POWSIM (Ryman & Palm, 2006). The default Markov chain parameters were used with the overall combined allele frequencies, dividing them into two populations with effective population size (*Ne*) of 10,000 individuals, and for varying number of generations (proportional to *Fst*) and sample sizes of 100 and 50 individuals. The software CREATE was used to produce input files whenever possible (Coombs *et al.*, 2008). Observed (H_O) and expected heterozygosity (H_E) were estimated with GENALEX (Peakall & Smouse, 2001) and GENEPOP V4.0 (Rousset, 2008). The same software was used to check genotypes for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium. Markov Chain parameters for the HWE test in GENEPOP V4.0 were 10,000 dememorisations and 100 batches of 5,000 iterations.

Table 6.1: Sample information.

Location= name of location where the sample was collected; Abv= the abreviation used in the text to refer to samples collected in a particular location, usually followed by the year the sample was collected (i.e. NeD06); Area= the general Area where the location is situated: NS=North Sea, EC= English Channel, ISS = Irish Sea South, ISN = Irish Sea North, AT= Atlantic coast of Ireland; Lat= Latitude; Long= Longitude; Sample Size = number of individual fish within a sample, sorted by collection year. Samples sizes after removal of incomplete genotypes are in *italics*; Ho and He= Observed and Expected Heterozygosities, sorted by sampling location and year; p-Value HWE (no Nulls)= Probability that the sample is in significant departure of Hardy-Weinberg Equilibrium expectations when marker with null alleles are removed. Significant P values (P<0.001) are in bold.

					Sample Size								*Ho* and *He*								P-Value HWE (no nulls)			
Location	**Abv.**	**Area**	**Lat.**	**Long.**	**2005**		**2006**		**2007**		**2008**		**2005**		**2006**		**2007**		**2008**		**2005**	**2006**	**2007**	**2008**
North East Dogger	**NeD**	**NS**	55.30	2.90	81	*64*	92	*87*	183	*180*	92	*72*	0.689	0.727	0.720	0.736	0.699	0.735	0.694	0.731	0.428	0.827	0.143	0.052
Amble	**Amb**	**NS**	55.27	-1.25	42	*34*	86	*77*	47	*44*	95	*85*	0.705	0.729	0.687	0.722	0.699	0.734	0.728	0.740	0.448	0.812	0.505	0.367
Off Flamborough	**Off**	**NS**	54.25	0.50	21	*17*	37	*19*	48	*47*	94	*91*	0.714	0.724	0.666	0.725	0.716	0.727	0.709	0.736	0.688	0.294	0.205	0.313
Indefatigable Bank	**Inf**	**NS**	53.56	2.08	-		59	*46*	48	*48*	100	*96*	-	-	0.701	0.731	0.704	0.721	0.690	0.735	-	0.274	0.721	0.324
Rye Bay	**Rye**	**EC**	50.78	0.73	-		92	*91*	178	*174*	96	*80*	-	-	0.716	0.729	0.702	0.734	0.680	0.726	-	0.895	0.092	0.451
Lyme Bay	**Lyb**	**EC**	50.61	-2.93	46	*41*	77	*69*	50	*50*	50	*42*	0.694	0.721	0.686	0.726	0.691	0.742	0.703	0.738	0.078	0.614	0.260	0.849
South Celtic Sea	**Ces**	**ISS**	*51.32*	*-7.46*	-		94	*79*	-		-		-	-	0.705	0.733	-	-	-	-	-	0.852	-	-
South Cardigan Bay	**Scb**	**ISS**	52.18	-4.50	-		99	*98*	-		-		-	-	0.706	0.732	-	-	-	-	-	0.159	-	-
Inner Cardigan Bay	**Inc**	**ISS**	52.30	-4.27	99	*79*	39	*36*	49	*47*	30	*30*	0.684	0.729	0.671	0.723	0.679	0.716	0.680	0.718	0.764	0.321	0.091	0.896
Red Wharf Bay	**Rwb**	**ISN**	53.74	-4.18	-		50	*50*	50	*50*	-		-	-	0.685	0.722	0.698	0.710	-	-	-	0.128	0.952	-
Liverpool Bay	**Liv**	**ISN**	53.47	-3.70	-		47	*42*	183	*173*	95	*41*	-	-	0.705	0.739	0.704	0.732	0.670	0.730	-	0.529	0.781	0.301
St. Bees Point	**Stb**	**ISN**	54.51	-3.79	30	*18*	95	*92*	48	*48*	96	*77*	0.667	0.712	0.679	0.718	0.684	0.713	0.676	0.727	0.751	0.534	0.135	0.300
Dundrum Bay	**Dub**	**ISN**	54.08	-5.62	-		64	*63*	-		-		-	-	0.724	0.740	-	-	-	-	-	0.186	-	-
North of Ireland	**Noi**	**AT**	55.87	-7.45	-		96	*83*	-		-		-	-	0.694	0.726	-	-	-	-	-	0.960	-	-
West of Ireland	**Woi**	**AT**	51.11	-11.25	-		48	*28*	-		-		-	-	0.705	0.724	-	-	-	-	-	0.977	-	-
Overall		-			319	*253*	1075	*960*	884	*861*	748	*614*									**<0.001**	**<0.001**	**<0.001**	**<0.001**

The presence of null alleles was assessed with MICRO-CHECKER (van Oosterhout *et al.*, 2004). Whenever samples or loci were not in HWE, and linkage disequilibrium or null alleles were detected, the source was identified and the electropherogram re-checked for veracity.

6.3.2.2 Single locus information content:

Each locus has a unique evolutionary history depending on when the locus became polymorphic within the genome, the locus and allele-specific mutation rates, and the population processes shaping the distribution of alleles. Therefore, the information content regarding population structure will also be unique to each locus. The information content of a locus, and thus its value from a population genetics perspective, will be assessed here by the amount of differentiation it is able to detect in the studied system.

The following estimators of genetic differentiation were calculated for each locus independently: Θ_{WC} and Θ'_{WC}, G_{ST_est} and G_{ST_est}', φ_{ST} and φ'_{ST}, D_S and D_{est}. Locus-specific Θ_{WC} (Weir & Cockerham, 1984) was calculated in GENEPOP V4.0 (Rousset, 2008). Standardised Θ'_{WC} estimates were calculated in a similar way to G'_{ST} (Hedrick, 2005). Θ_{max} was calculated by transforming the raw genotype data with RECODEDATA V.0.1 (Meirmans, 2006) so that all populations had non-overlapping allele ranges for all loci. In these conditions, all populations are maximally differentiated (no alleles in common), but heterozygosities remain unaffected by the transformation. Single locus Θ_{max} was calculated also with GENEPOP V4.0 (Rousset, 2008). Standardized Θ'_{WC} was calculated by dividing the original Θ_{WC} by Θ_{max}. Other studies have recently applied similar approaches to calculate HCEs of F_{ST} (Kenchington *et al.*, 2009; Papetti *et al.*, 2009). Single locus φ_{ST} (Excoffier *et al.*, 1992) and its heterozygosity corrected standardised version, φ'_{ST} (Meirmans, 2006), were calculated in GENODIVE (Meirmans & van Tiernderen, 2004) with significance values based on 10,000 permutations. G'_{ST} and D were calculated manually following published instructions (Hedrick, 2005; Jost, 2008) and confirmed with the software SMoG-D (Crawford, 2009), which was also used to calculate G'_{ST_est} and D_{est} which are corrected for differences in sample size (Nei & Chesser, 1983). To assess the influence of the correction in each locus, the magnitude of the increase (Δ_{HC}) from UCEs to HCEs was calculated for one of the estimators (Θ_{WC}) by subtracting the UCE from the HCE (i.e. $\Delta_{HC} = \Theta'_{WC} - \Theta_{WC}$) for each locus and year.

Investigation of the HTEs was undertaken using two versions of the dab data set: data set A consisted of the entire data; while data set B consisted of only fully genotyped individuals (89% of the entire data set) and excluded loci with evidence of null alleles, allowing an examination of the influence of missing data and loci with null alleles. The relationships

(Pearson's correlation and associated probabilities) between the different estimators of differentiation and H_E or probabilities of random genic and genotypic distributions were analysed using MINITAB®.

Two additional contrasting analysis methods were employed to evaluate the relevance of the heterozygosity correction on a locus by locus basis. First, a locus specific Exact *G*-test (Raymond & Rousset, 1995), which reports the probability that the alleles observed at a locus are randomly distributed among samples (i.e. the locus is not structured). Exact *G*-tests were calculated in GENEPOP V4.0 (Rousset, 2008) with 10,000 dememorisations and 500 batches of 10,000 iterations. And secondly, a multivariate analysis, in which the allelic patterns provided by each loci can be studied independently and contrasted (Laloë *et al.*, 2007; Jombart *et al.*, 2009). The consensus of genetic structuring of the different markers was assessed by correspondence analysis of single locus population data. The analysis was performed with the packages ADE4 (Chessel *et al.*, 2004) and ADEGENET (Jombart, 2008) in R (R Development Core Team, 2009), to evaluate the magnitude, the direction, and the temporal stability of the information portrayed by each locus. Missing allelic data were substituted by the mean χ^2 distance, which effectively places missing data at the origin of the axis (Jombart, 2008). Using genetic markers under directional or balancing selection can distort estimates of genetic differentiation (Nielsen *et al.*, 2006), therefore the neutrality of markers was tested with the F_{ST}-outlier method (Beaumont & Nichols, 1996; Beaumont & Balding, 2004) in LOSITAN (Antao *et al.*, 2008) for each year independently and the whole data set combined.

6.3.2.3 Multilocus pairwise population differentiation:

Locus-specific pairwise Θ_{WC} (Weir & Cockerham, 1984), Θ_{max}, Θ'_{WC}, and pairwise Exact *G*-test of genic subdivision (Raymond & Rousset, 1995) were calculated in GENEPOP 4.0. Pairwise Θ_{WC}, Nei's D_S (Nei, 1972) and unbiased D_{SU} (Nei, 1978), and associated significances based on 1000 permutations, were calculated with GENETIX (Belkhir *et al.*, 1996-2004). Pairwise φ'_{ST} and significance values based on 10,000 permutation were calculated in GENODIVE (Meirmans & van Tiernderen, 2004). Pairwise D_{est} and G'_{ST_est} were computed in SMoG-D (Crawford, 2009). As recommended by Rousset (2008), multilocus averages were calculated for Θ_{WC} in GENEPOP 4.0. However, another multilocus average (arithmetic mean of all loci) was calculated for Θ_{WC} and for all other estimators so they could be compared. For all estimators, each year was treated independently as suggested by Balloux & Lugon-Moulin (2002). Loci with null alleles were excluded from the multilocus average calculation.

6.4 Results:

6.4.1 Power, locus characteristics and HWE:

POWSIM suggested that the power of the loci combination was 99% to detect differentiations as low as 0.0025 with sample sizes of 100 individuals, or as low as 0.005 for sample sizes of 50 individuals. The genotyping error rate was low with 98.4% of alleles identical in both amplifications of the duplicated sample, and most of the errors were at two microsatellite loci: *DAC1-35* (92% accuracy) and *DAC5-70* (96% accuracy). The number of alleles per locus ranged between 8 and 56, and the observed heterozygosity per locus was between 0.089 and 0.946 (Table 6.2). The mean observed heterozygosity across loci within samples was between 0.666 and 0.728. The exclusion of *DAC1-35* and *DAC5-70* did not alter significantly the mean observed heterozygosity. Two microsatellites, *DAC1-35* and *DAC5-70*, showed deviations from HWE in the form of heterozygote deficiency (Table 6.2). MICRO-CHECKER suggested the presence of null alleles in some of the samples in some of the years for *DAC1-35* and for all samples in all years for *DAC5-70*. No markers showed evidence of large allele drop-out or stuttering scoring problems. Once the two markers with heterozygosity deficiencies were removed, all samples conformed to HWE expectations. When all samples were pooled into a single group each year, HW became highly significant, even after removal of *DAC1-35* and *DAC5-70*, suggesting that all samples do not belong to the same panmictic population. No linkage disequilibrium was found among any combination of loci when all samples were pooled together. However, significant linkage ($p<0.001$) was found in up to three genotype comparisons in each year, but none were consistent across samples or years. The source of the disequilibrium was always one or two individuals showing rare alleles or unusual allele combinations at two or more loci for the particular sample collected.

6.4.2 Locus neutrality test:

The neutrality test conducted using LOSITAN suggested that *DAC3-14* was outside the 95% confidence envelope for each of the four years and the whole data set, suggesting that directional selection may have played a role in shaping the allele distribution at that locus. Although one locus (16x0.05=0.8) would be predicted to be significant by chance alone, the annual recurrence deserves attention. Four other loci (*DAG2-90*, *DAG5-17*, *DAC5-21*, and *DAC1-35*) appeared to be under balancing selection at 95% CI when the whole data set was analysed together, but not on a year by year basis, when the CI was set to 99%, or when *DAC3-14* was removed from the analysis.

Table 6.2: Locus information

Locus= locus name; Na = Number of alleles at locus; Range = Range of allele sizes (bp); $\hat{H}o$ = Observed heterozygosity at locus; $\hat{H}e$= Mean expected heterozygosity; p-value of HWE conformity= probability that the locus is in Hardy-Weinberg equilibrium expectations, significant values (p<0.001) are in bold; P-value of Genic differentiation = probability that the allele diversity found at the locus is randomly distributed across samples, values significantly departing from random distribution (p<0.001) are in bold; Fis (GENEPOP) = locus-specific Fis as calculated by GENEPOP.

	Locus	DAC1-35	DAC1-55	DAC2-15	DAC2-28	DAC5-5	DAG2-90	DAG4-64	DAC3-14	DAC4-40	DAC5-77	DAG5-17	DAC1-90	DAC2-37	DAC5-21	DAC5-70	DAG5-88	Mean
Basic statistics	Na	56	22	25	25	15	23	34	18	30	13	51	30	8	43	55	9	**28.6**
	Range (bp)	300-416	234-276	186-224	108-156	95-129	156-200	134-209	150-192	312-368	98-124	154-260	102-162	238-352	91-176	98-205	179-197	-
	$\hat{H}o$	0.804	0.815	0.773	0.809	0.196	0.907	0.925	0.687	0.884	0.460	0.946	0.911	0.089	0.927	0.580	0.409	
	$\hat{H}e$	0.942	0.824	0.761	0.812	0.195	0.904	0.947	0.768	0.884	0.461	0.964	0.941	0.094	0.944	0.939	0.416	
P-Value of HWE conformat	2005	**<0.001**	0.0753	0.8805	0.2246	0.8735	0.7447	0.6726	0.2305	0.2568	0.8615	0.4659	0.2577	0.8426	0.4243	**<0.001**	0.3118	
	2006	**<0.001**	0.8219	0.5011	0.4518	0.3681	0.235	0.1541	0.173	0.9161	0.9008	0.7306	0.6887	0.4505	0.483	**<0.001**	0.8993	
	2007	**<0.001**	0.3467	0.6387	0.4689	0.2562	0.5958	0.7561	0.0031	0.4641	0.6935	0.5699	0.182	0.2159	0.4524	**<0.001**	0.2393	
	2008	**<0.001**	0.1677	0.1546	0.5952	0.9003	0.1875	0.6246	0.1392	0.1453	0.9401	0.1562	0.5805	0.8626	0.2268	**<0.001**	0.5141	
P-Value of Genic Differentia	2005	0.109	0.211	0.261	0.446	0.208	0.244	**0.000**	**0.000**	**0.000**	0.651	0.440	**0.001**	0.367	**0.002**	**0.002**	0.637	**<0.001**
	2006	0.109	**0.000**	0.023	0.480	**0.000**	0.041	0.013	**0.000**	**0.000**	0.362	0.013	**0.001**	**0.000**	0.097	**0.000**	0.532	**<0.001**
	2007	**0.000**	0.036	0.012	0.779	0.080	0.574	**0.000**	**0.000**	**0.000**	0.108	0.023	0.019	0.009	0.091	**0.000**	0.798	**<0.001**
	2008	0.004	**0.000**	0.012	0.095	**0.001**	0.721	**0.000**	**0.000**	**0.000**	0.325	0.189	**0.000**	0.015	0.017	**0.000**	0.133	**<0.001**
f_{wc} (GENEPOP)	2005	0.110	0.004	0.000	0.005	0.016	-0.002	0.011	0.078	0.015	-0.006	0.020	0.058	0.060	0.019	0.445	-0.051	0.059
	2006	0.134	0.011	-0.014	-0.002	0.045	-0.006	0.020	0.061	-0.009	-0.010	0.014	0.017	0.071	0.019	0.342	0.011	0.048
	2007	0.119	0.002	-0.011	0.021	0.040	-0.011	0.003	0.101	0.001	-0.006	0.005	0.016	0.090	0.004	0.356	-0.002	0.048
	2008	0.183	0.007	-0.037	0.002	0.009	0.007	0.037	0.060	0.004	-0.013	0.022	0.026	0.024	0.025	0.401	0.052	0.060
Θ_{wc}	2005	-0.001	0.003	0.002	0.000	0.002	0.001	0.008	0.025	0.010	-0.001	0.000	0.005	0.003	0.002	0.000	0.007	
	2006	-0.001	0.004	0.001	0.000	0.008	0.001	0.002	0.022	0.005	0.000	0.001	0.002	0.011	0.001	0.003	0.000	
	2007	0.002	0.004	0.003	0.000	0.004	-0.001	0.004	0.027	0.003	0.000	0.001	0.000	0.009	0.001	0.004	-0.003	
	2008	0.000	0.006	0.005	0.002	0.017	0.000	0.005	0.015	0.004	0.001	0.000	0.003	0.011	0.002	0.003	0.008	
Θ'_{wc}	2005	-0.009	0.016	0.007	0.000	0.003	0.009	0.133	0.103	0.082	-0.001	-0.014	0.076	0.003	0.027	0.007	0.011	
	2006	-0.009	0.023	0.006	0.000	0.010	0.014	0.036	0.087	0.043	0.000	0.034	0.035	0.012	0.020	0.046	0.000	
	2007	0.031	0.020	0.014	-0.002	0.005	-0.006	0.069	0.109	0.023	0.000	0.036	0.008	0.010	0.012	0.061	-0.006	
	2008	-0.004	0.033	0.020	0.009	0.021	-0.004	0.082	0.062	0.035	0.002	0.005	0.045	0.012	0.031	0.045	0.015	

6.4.3 Single locus information content:

6.4.3.1 Estimators of differentiation:

The locus-specific UCEs (Θ_{WC}, G_{ST_est}, and φ_{ST}) ranged from negative and nil (*DAC1-35*, *DAC2-28*, *DAC5-77*, *DAG5-88*) to 0.027 (*DAC3-14*), indicating a range of differentiation values across the loci. HTEs (Θ'_{WC}, G'_{ST_est}, φ'_{ST}, and D_{est}) also had minimum values of zero (*DAC2-28*, *DAC5-77*, and *DAG2-90* and *DAG5-88* in most years), while maximum values were much higher for some loci (*DAC3-14*, *DAG4-64*, *DAC4-40*, and *DAC1-90*) and reached up to 0.100 in some cases (Table 6.3). Predictably, locus-specific Θ_{max} had a linear inversely proportional relationship with heterozygosity (Figure 6.1), and was equal to the expected homozygosity (Θ_{max}=1-H_E), thus low heterozygosity markers (*DAC2-37, DAC5-5, DAG5-88* and *DAC5-77*) exhibited the greatest potential to reveal differentiation as measured by UCEs (up to 0.900), while high heterozygosity markers (*DAG5-17, DAC5-21, DAG4-64* and *DAC1-90*) showed Θ_{max} values around 0.050 (the reader is reminded that this in complete absence of common alleles). Accordingly, Θ_{WC} also exhibited the same trend, with higher values for low diversity markers (*DAC5-5* and *DAC2-37*) (Figure 6.2). The standardisation of Θ_{WC} markedly changed the perception of information content of the different loci. For example, highly heterozygous *DAG4-64* increased from a Θ_{WC} of 0.003 to a Θ'_{WC} of 0.080, while the relative contribution to multilocus differentiation of low variability *DAC2-37* declined. Some markers did not change in ranking: *DAC3-14* ranked highly with all estimators, while *DAG2-90* remained non-informative despite containing high gene diversity. The relationship between HTE and H_E was clearly positive (Figure 6.3; Figure 6.4). The values of genetic differentiation obtained by the different HTE were very similar to each other, as were those from uncorrected estimates, with the exception of those loci known to have null alleles (*DAC1-35* and *DAC5-70*). Estimates of loci with null alleles based on Θ_{WC} and φ_{ST} showed reduced differentiation compared to estimates based on G_{ST_est} and D_{est}, perhaps indicating that the latter may be prone to overestimation of differentiation when null alleles are present.

Table 6.3: Locus-specific differentiation values (information content)

Locus-specific differentiation values (information content) of 16 microsatellite loci in dab, *Limanda limanda*, for four years (2005-2008). **Locus**= locus name; H_E= Heterozygosity (colour coded from blue=low to green=high); Uncorrected estimators of differentiation = Θ_{WC}, G_{ST_est} , and φ_{ST} ; Heterozygosity corrected estimators = Θ'_{WC}, G_{ST_est}', φ'_{ST}, D_{est} (All colour coded from yellow=low to orange=high); **p(gnc)** = probability that the allele distribution at the locus is random among sample; **p(gtc)** = probability that the genotype distribution is random among samples (p significant after Bonferroni correction are in **bold**); Δ_{HC} = magnitude of change after heterozygosity correction ($\Delta_{HC} = \Theta'_{WC} - \Theta_{WC}$).

	Locus	DAC1-35	DAC1-55	DAC2-15	DAC2-28	DAC5-5	DAG2-90	DAG4-64	DAC3-14	DAC4-40	DAC5-77	DAG5-17	DAC1-90	DAC2-37	DAC5-21	DAC5-70	DAG5-88
2005	H_E	0.942	0.818	0.773	0.804	0.220	0.903	0.942	0.756	0.875	0.438	0.965	0.939	0.073	0.943	0.946	0.395
	Θ_{WC}	-0.001	0.003	0.002	0.000	0.002	0.001	0.008	0.025	0.010	-0.001	0.000	0.005	0.003	0.002	0.000	0.007
	G_{ST_est}	0.000	0.000	0.001	0.001	0.000	0.001	0.005	0.024	0.011	0.000	-0.001	0.005	-0.002	0.002	0.006	0.013
	φ_{ST}	0.000	0.002	0.001	0.003	-0.002	0.005	0.007	0.033	0.009	0.008	-0.001	0.005	0.002	0.001	-0.002	0.000
	Θ'_{WC}	-0.009	0.016	0.007	0.000	0.003	0.009	0.133	0.103	0.082	-0.001	-0.014	0.076	0.003	0.027	0.007	0.011
	G'_{ST_est}	0.000	0.001	0.006	0.004	0.000	0.011	0.113	0.111	0.102	0.000	-0.020	0.099	-0.002	0.038	0.112	0.024
	φ'_{ST}	0.007	0.012	0.005	0.013	-0.003	0.051	0.122	0.137	0.071	0.014	-0.036	0.075	0.002	0.017	-0.037	-0.001
	D_{est}	0.000	0.001	0.005	0.003	0.000	0.010	0.108	0.089	0.092	0.000	-0.019	0.094	0.000	0.037	0.107	0.010
	p (gnc)	0.098	0.204	0.280	0.440	0.202	0.236	**0.000**	**0.000**	**0.000**	0.652	0.456	**0.001**	0.373	**0.002**	**0.003**	0.626
	p (gtp)	0.387	0.270	0.260	0.451	0.243	0.252	**0.000**	**0.000**	**0.000**	0.689	0.535	**0.002**	0.414	**0.003**	0.737	0.600
	Δ_{HC}	-0.008	0.013	0.006	0.000	0.001	0.008	0.125	0.078	0.071	-0.001	-0.013	0.071	0.000	0.025	0.006	0.004
2006	H_E	0.942	0.828	0.750	0.809	0.213	0.902	0.945	0.748	0.877	0.456	0.963	0.940	0.091	0.943	0.941	0.398
	Θ_{WC}	-0.001	0.004	0.001	0.000	0.008	0.001	0.002	0.022	0.005	0.000	0.001	0.002	0.011	0.001	0.003	0.000
	G_{ST_est}	0.001	0.003	0.001	0.000	0.008	0.001	0.003	0.022	0.004	0.000	0.001	0.003	0.014	0.002	0.006	-0.002
	φ_{ST}	-0.001	0.004	0.000	-0.001	0.009	0.001	0.002	0.020	0.006	0.001	0.001	0.002	0.009	0.001	0.002	0.000
	Θ'_{WC}	-0.009	0.023	0.006	0.000	0.010	0.014	0.036	0.087	0.043	0.000	0.034	0.035	0.012	0.020	0.046	0.000
	G'_{ST_est}	0.009	0.021	0.006	0.000	0.011	0.016	0.050	0.092	0.036	0.000	0.038	0.056	0.016	0.030	0.094	-0.003
	φ'_{ST}	-0.012	0.023	0.001	-0.005	0.011	0.007	0.034	0.081	0.048	0.002	0.037	0.037	0.010	0.024	0.039	-0.001
	D_{est}	0.009	0.018	0.005	0.000	0.002	0.014	0.048	0.072	0.032	0.000	0.036	0.053	0.002	0.029	0.089	-0.001
	p (gnc)	0.121	**0.000**	0.027	0.527	**0.000**	0.038	**0.010**	**0.000**	**0.000**	0.363	**0.009**	**0.001**	**0.000**	0.097	**0.000**	0.539
	p (gtp)	0.831	**0.000**	0.033	0.494	**0.001**	0.028	0.025	**0.000**	**0.001**	0.347	0.024	0.004	**0.000**	0.217	**0.000**	0.547
	Δ_{HC}	-0.008	0.019	0.004	0.000	0.002	0.012	0.034	0.065	0.037	0.000	0.033	0.033	0.001	0.019	0.043	0.000

Table 6.3: (cont.)

	Locus	DAC1-35	DAC1-55	DAC2-15	DAC2-28	DAC5-5	DAG2-90	DAG4-64	DAC3-14	DAC4-40	DAC5-77	DAG5-17	DAC1-90	DAC2-37	DAC5-21	DAC5-70	DAG5-88
2007	H_E	0.945	0.820	0.756	0.816	0.166	0.906	0.944	0.749	0.881	0.457	0.964	0.940	0.099	0.943	0.932	0.428
	Θ_{WC}	0.002	0.004	0.003	0.000	0.004	-0.001	0.004	0.027	0.003	0.000	0.001	0.000	0.009	0.001	0.004	-0.003
	G_{ST_est}	0.002	0.003	0.002	-0.001	0.006	0.000	0.003	0.034	0.002	0.001	0.002	0.000	0.012	0.000	0.009	-0.004
	φ_{ST}	0.002	0.004	0.003	-0.001	0.005	-0.001	0.004	0.024	0.002	0.001	0.002	0.000	0.007	0.001	0.006	-0.004
	Θ'_{WC}	0.031	0.020	0.014	-0.002	0.005	-0.006	0.069	0.109	0.023	0.000	0.036	0.008	0.010	0.012	0.061	-0.006
	G'_{ST_est}	0.031	0.016	0.007	-0.007	0.008	-0.004	0.061	0.139	0.015	0.001	0.057	-0.001	0.014	0.010	0.127	-0.007
	φ'_{ST}	0.028	0.021	0.013	-0.003	0.005	-0.012	0.068	0.097	0.019	0.003	0.046	0.002	0.008	0.015	0.084	-0.006
	D_{est}	0.029	0.013	0.005	-0.006	0.001	-0.003	0.058	0.110	0.013	0.001	0.055	-0.001	0.002	0.009	0.119	-0.003
	p (gnc)	**0.000**	0.030	0.016	0.763	0.088	0.570	**0.000**	**0.000**	**0.000**	0.099	0.027	0.017	0.009	0.073	**0.000**	0.762
	p (gtp)	0.009	0.038	0.014	0.792	0.138	0.538	**0.000**	**0.000**	**0.000**	0.087	0.025	0.026	0.018	0.114	**0.004**	0.888
	Δ_{HC}	0.029	0.017	0.010	-0.002	0.001	-0.005	0.065	0.082	0.020	0.000	0.035	0.007	0.001	0.011	0.057	-0.003
2008	H_E	0.939	0.808	0.769	0.809	0.186	0.905	0.943	0.754	0.885	0.482	0.964	0.938	0.106	0.943	0.930	0.432
	Θ_{WC}	0.000	0.006	0.005	0.002	0.017	0.000	0.005	0.015	0.004	0.001	0.000	0.003	0.011	0.002	0.003	0.008
	G_{ST_est}	0.001	0.007	0.004	0.001	0.023	-0.001	0.004	0.016	0.003	0.000	0.001	0.002	0.012	0.002	0.006	0.008
	φ_{ST}	0.001	0.007	0.006	0.002	0.020	0.000	0.004	0.010	0.004	0.002	0.001	0.002	0.014	0.001	0.003	0.011
	Θ'_{WC}	-0.004	0.033	0.020	0.009	0.021	-0.004	0.082	0.062	0.035	0.002	0.005	0.045	0.012	0.031	0.045	0.015
	G'_{ST_est}	0.013	0.041	0.018	0.003	0.029	-0.011	0.078	0.070	0.030	0.000	0.025	0.042	0.014	0.037	0.091	0.015
	φ'_{ST}	0.014	0.038	0.026	0.011	0.024	-0.005	0.070	0.042	0.037	0.003	0.018	0.037	0.015	0.019	0.041	0.019
	D_{est}	0.013	0.034	0.014	0.003	0.007	-0.010	0.074	0.055	0.027	0.000	0.024	0.040	0.002	0.035	0.086	0.007
	p (gnc)	**0.003**	**0.000**	0.010	0.094	**0.001**	0.744	**0.000**	**0.000**	**0.000**	0.337	0.180	**0.000**	0.017	0.019	**0.000**	0.132
	p (gtp)	0.221	**0.000**	0.009	0.100	**0.001**	0.741	**0.000**	**0.000**	**0.000**	0.360	0.242	**0.000**	0.032	0.041	0.014	0.155
	Δ_{HC}	-0.004	0.027	0.015	0.007	0.004	-0.004	0.077	0.047	0.031	0.001	0.005	0.042	0.001	0.029	0.042	0.006

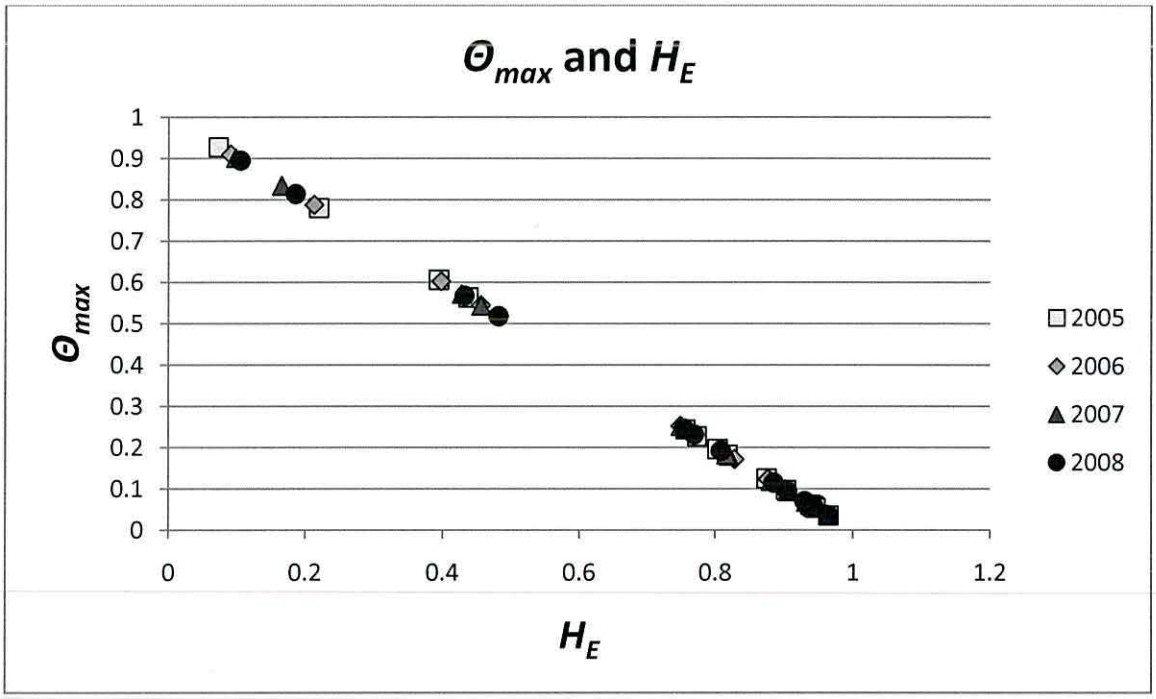


Figure 6.1: Relationship between Θ_{max} and H_E.
Relationship between locus-specific Θ_{max} and H_E among dab samples for 16 loci. Each year is represented by a different symbol.

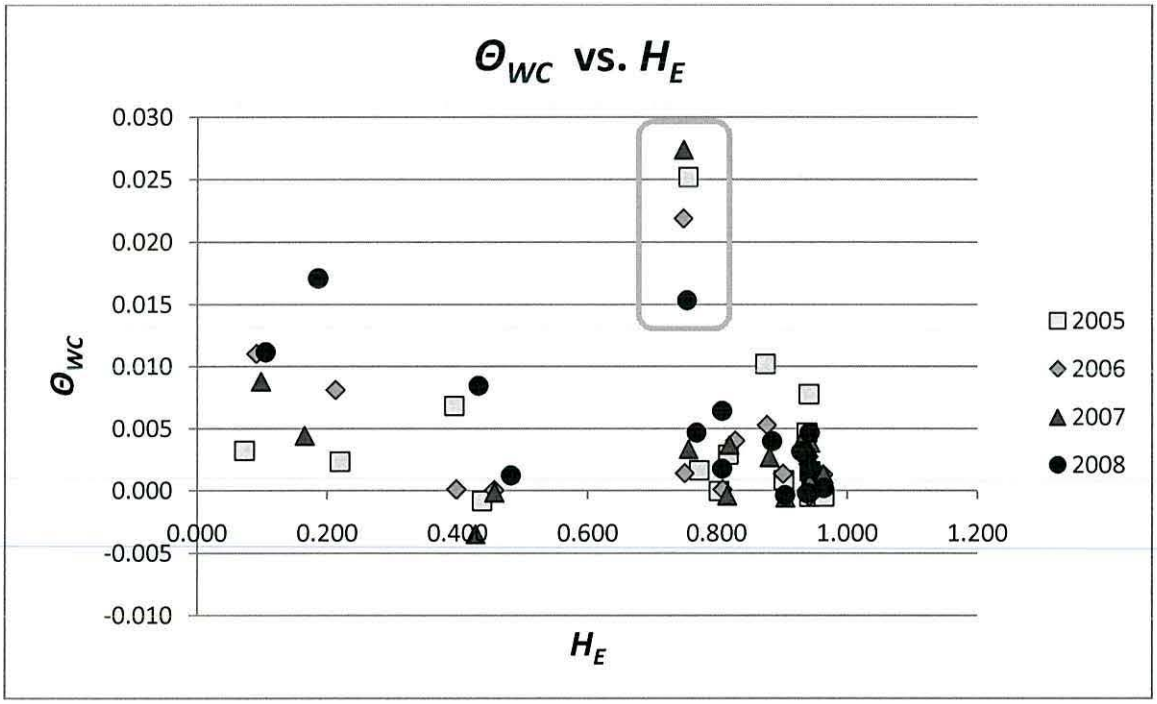


Figure 6.2: Relationship between Θ_{WC} and H_E.
Relationship between locus-specific Θ_{WC} and H_E among dab samples for 16 loci. Each year is represented by a different symbol. Locus *DAC3-14* is highlighted within the grey box.

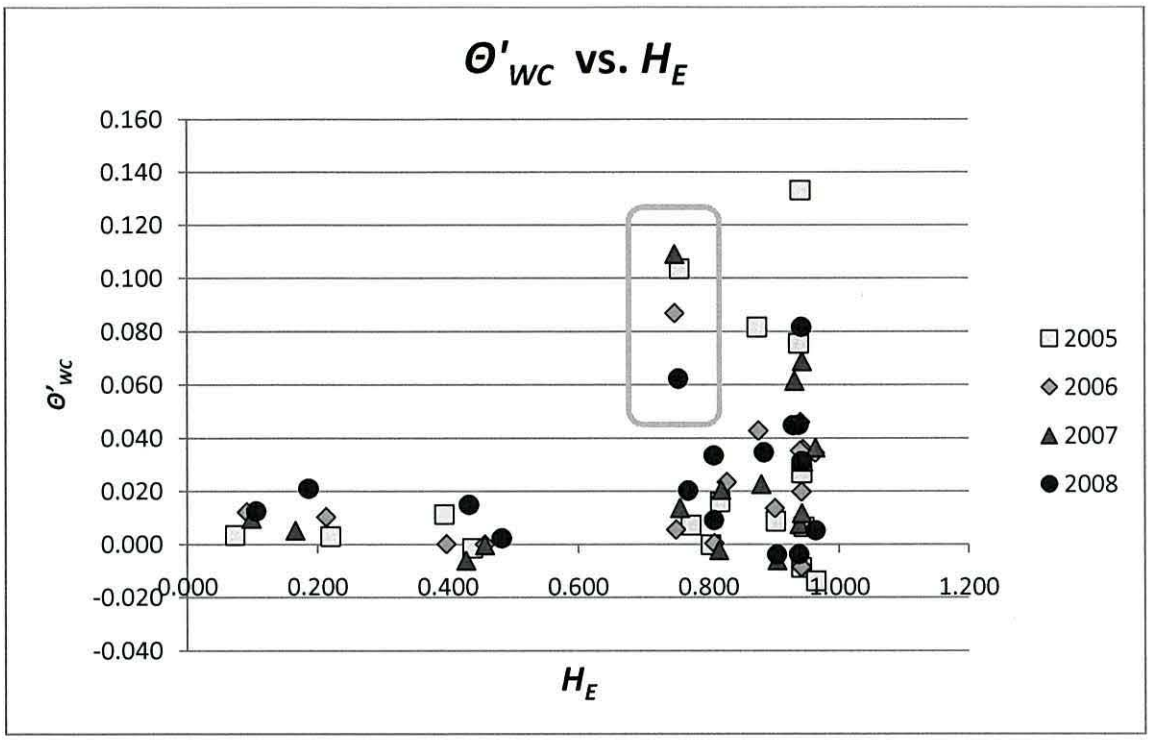


Figure 6.3: Relationship between Θ'_{WC} and H_E:
Relationship between locus-specific Θ'_{WC} and H_E among dab samples for 16 loci. Each year is represented by a different symbol. Locus *DAC3-14* is highlighted by the grey box.

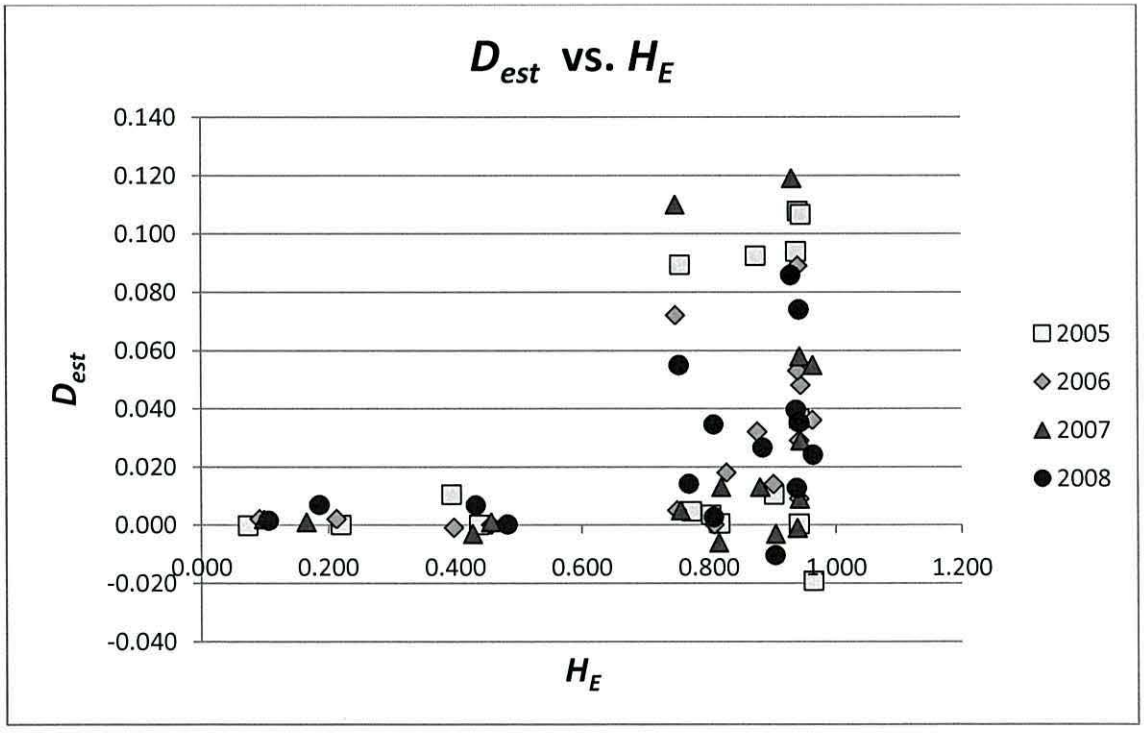


Figure 6.4: Relationship between D_{est} and H_E:
Relationship between locus-specific D_{est} and H_E among dab samples for 16 loci. Each year is represented by a different symbol.

6.4.3.2 Genic and Genotypic Exact G-test:

The analysis of genic and genotypic subdivision (i.e. whether alleles are randomly distributed across samples) revealed that allele frequencies were significantly different among samples ($p<0.001$) in all years for a number of loci (those with the highest Δ_{HC}: *DAG4-64*, *DAC3-14*, *DAC4-40*, *DAC5-70*), and significant ($p<0.05$) in at least two years for seven other loci (those with average Δ_{HC}: *DAC1-35*, *DAC1-55*, *DAC2-15*, *DAC5-5*, *DAG5-17*, *DAC2-37*, and *DAC5-21*), further suggesting the existence of genetic differentiation. No significant differences were detected in four markers (those with little or no Δ_{HC}: *DAC2-28*, *DAG2-90*, *DAC5-77*, and *DAG5-88*). When studied in more detail it became clear that there were marked and consistent differences in allele frequencies at some loci between some locations, particularly between basins (Figure 6.5).

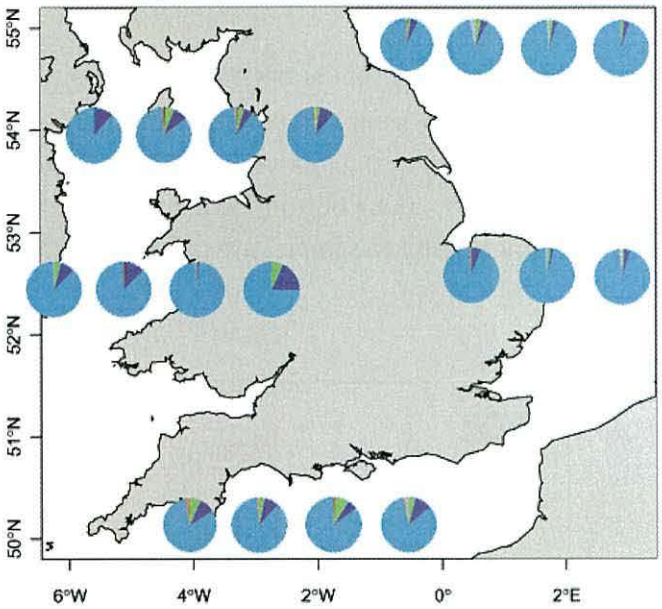
6.4.3.3 Multivariate Analysis:

The results of the single locus correspondence analysis (CA) showed that four loci (*DAC1-55*, *DAG4-64*, *DAC3-14*, and *DAC1-90*) separated North Sea and Irish Sea samples clearly along the first axis (The CA of these four loci are in Figure 6.6). These markers had the highest locus-specific Θ'_{WC}, while many had low Θ_{WC}, confirming that important information was missed by Θ_{WC}. Another three loci (*DAC2-15*, *DAC5-5*, and *DAC4-40*) also succeeded in separating North Sea and Irish Sea samples for some of the years, which coincided with those years with higher locus-specific Δ_{HC}. The remaining loci did not differentiate between North Sea and Irish Sea. These were the markers largely unaffected by the correction. Loci *DAG5-17* and *DAC5-21* are notable exceptions as, although the corrected values were relatively high, the CA did not suggest a division between North Sea and Irish Sea.

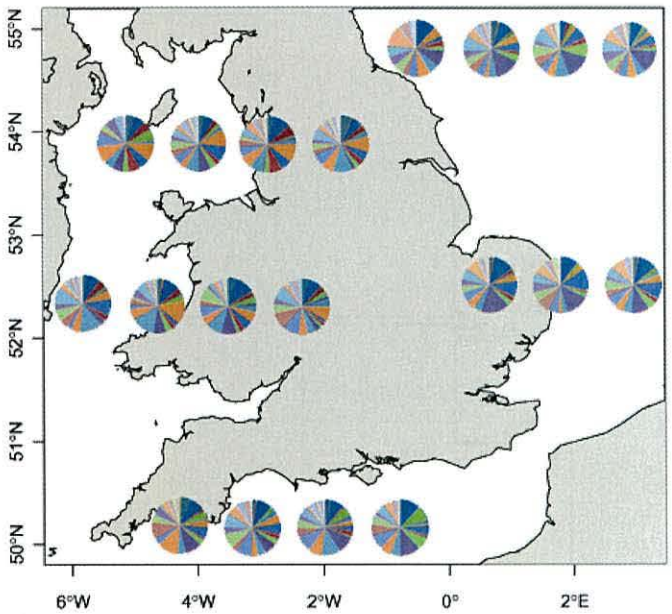
Figure 6.5: Selected allele frequency plots across the North Sea, English Channel and Irish Sea.

The pie charts depict the proportion of each allele in a sample. Each colour represents a different allele. There are four pie charts per area (three for InF) representing the different sampling years: 2005 to the left, 2008 to the right.

DAC5-5: Low diversity locus. Note the lower frequency of the main allele in blue (121bp) in LyB and the Irish Sea.



DAC4-64: High diversity locus. Note the higher frequency of the purple allele in the North Sea and LyB (165bp), and higher frequency of the orange allele in Irish Sea (157bp).



DAC3-14: Average diversity locus. Note the higher frequency of the red allele (168bp) in the North Sea, the higher frequency of the blue one in the Irish Sea (162bp), and intermediate frequency of both in LyB.

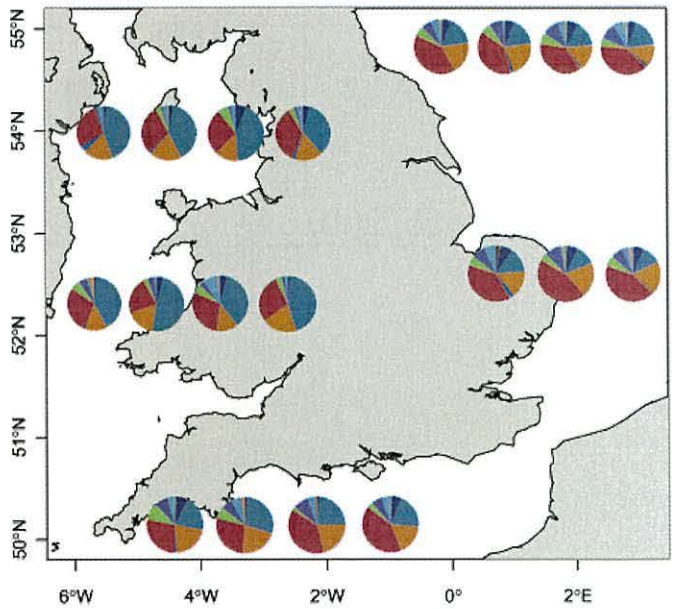
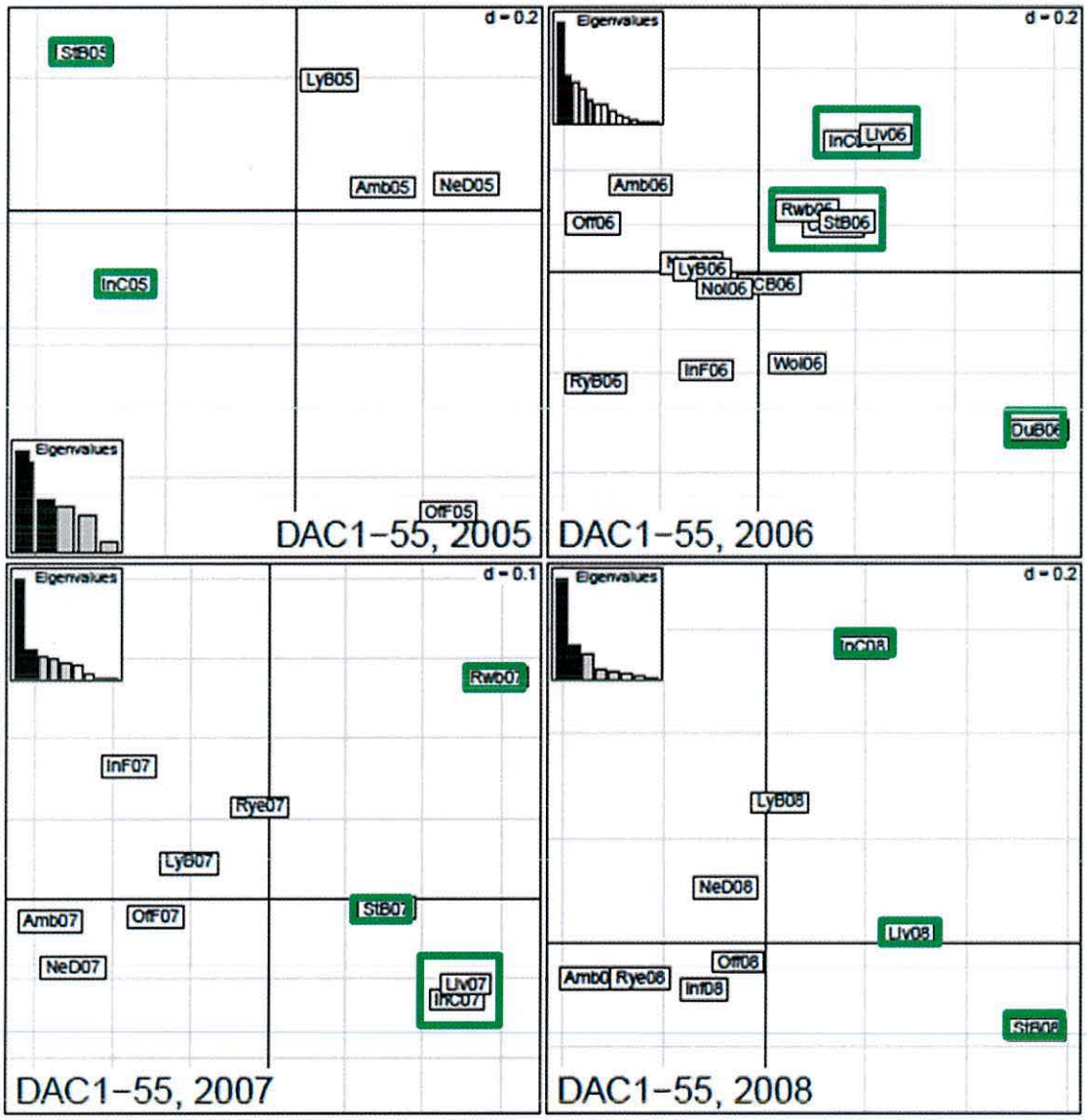
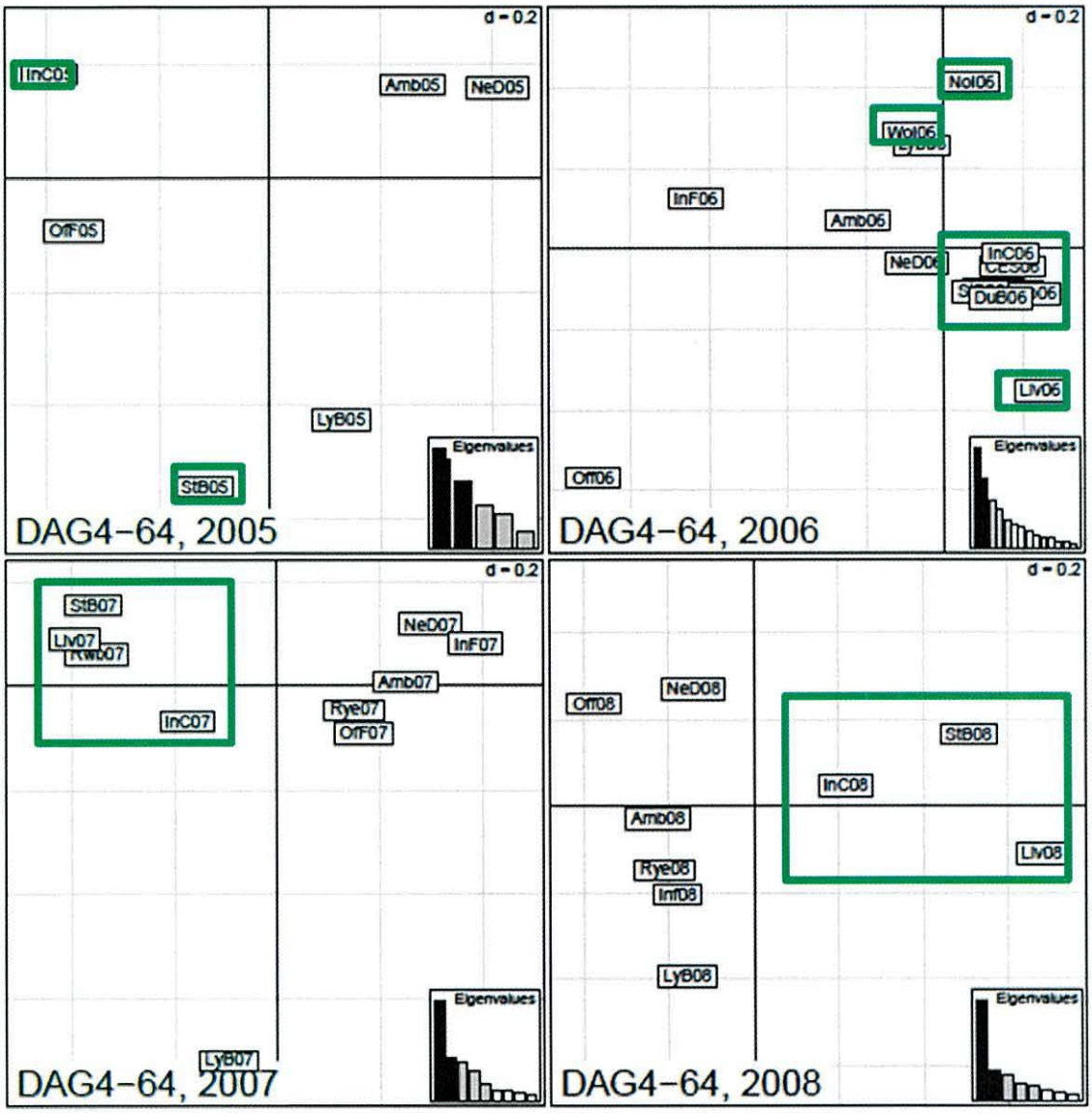
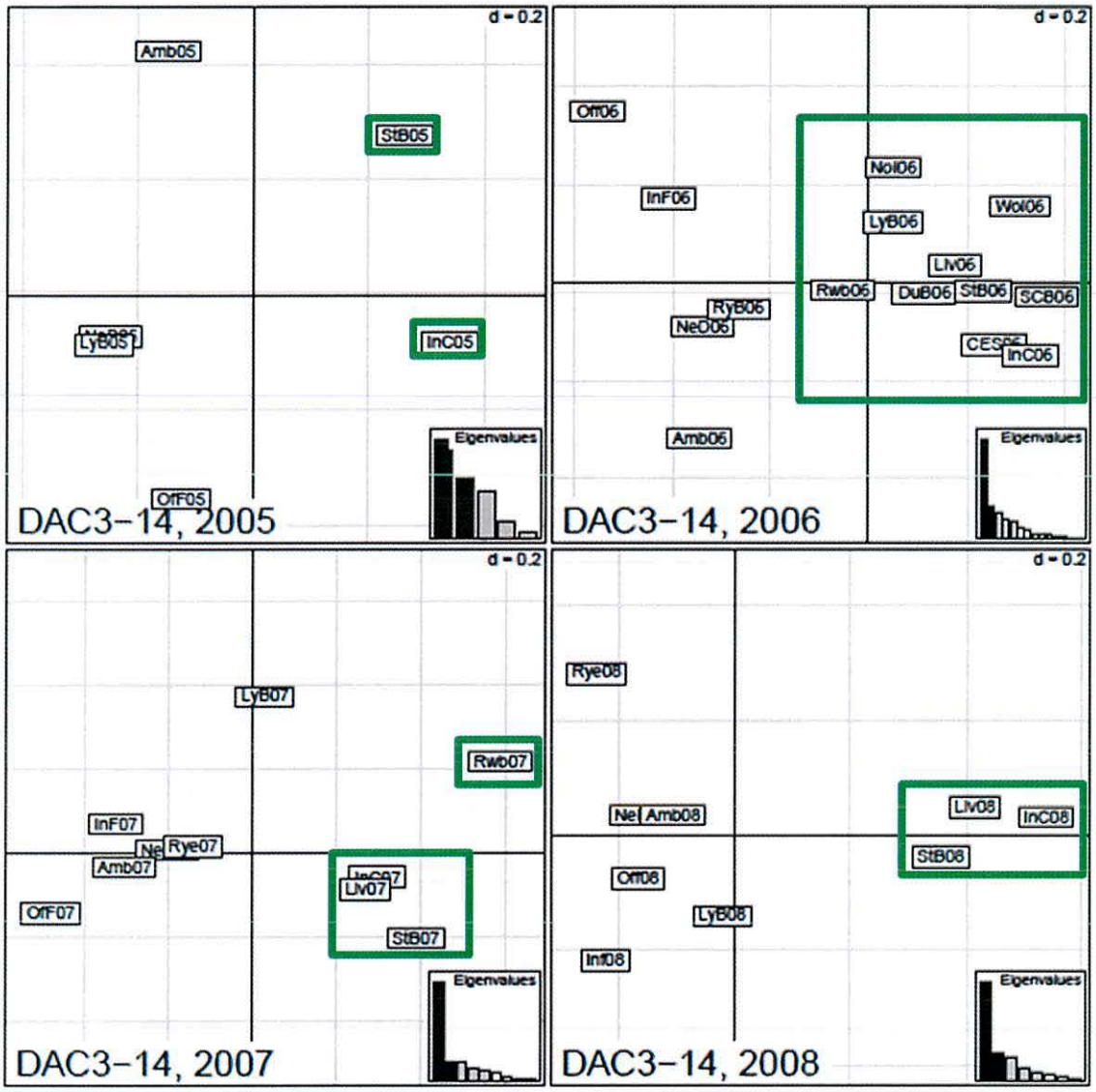


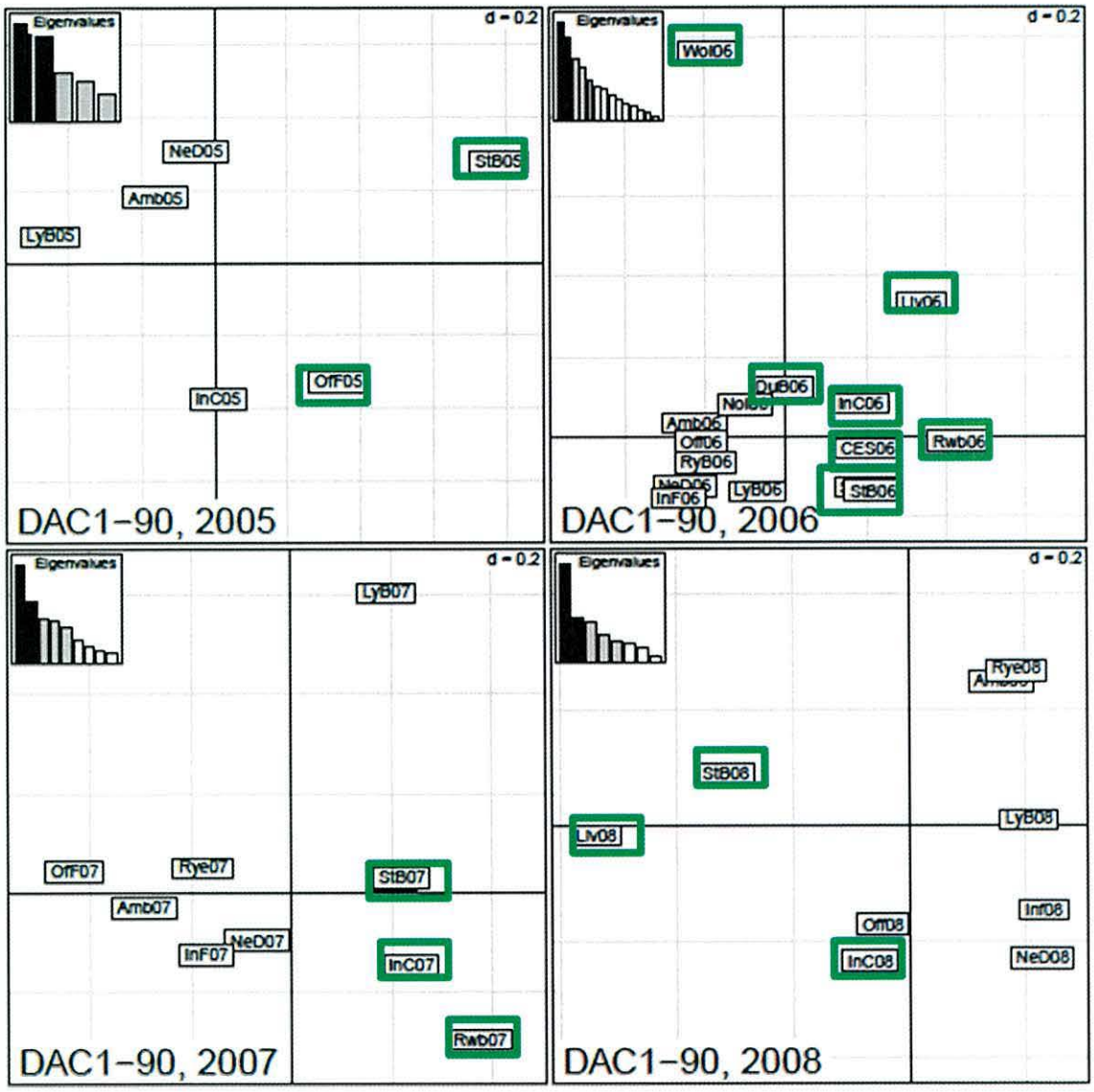
Figure 6.6: Locus-specific sample correspondence analysis

Locus-specific sample correspondence for four loci: *DAC1-55*, *DAG4-64*, *DAC3-14*, and *DAC1-90.* In each plot the four years are represented: 2005 - top left; 2006 - top right; 2007 - bottom left; 2008 - bottom right. In each plot, the screeplots of the eigenvalues are drawn in a corner. Irish Sea samples are coloured in greed to ease identification.




d = 0.2
InC05
Amb05
NeD05
OffF05
LyB05
StB05
Eigenvalues
DAG4-64, 2005
d = 0.2
Nol06
Wol06
InF06
Amb06
NeD06
InC06
DuB06
Liv06
Off06
Eigenvalues
DAG4-64, 2006
d = 0.2
StB07
Liv07
NeD07
InF07
Amb07
InC07
Rye07
OffF07
LyB07
Eigenvalues
DAG4-64, 2007
d = 0.2
Off08
NeD08
StB08
InC08
Amb08
Liv08
Rye08
InF08
LyB08
Eigenvalues
DAG4-64, 2008


d = 0.2
Amb05
StB05
LyB05
InC05
OffF05
Eigenvalues
DAC3-14, 2005
d = 0.2
Off06
NoI06
InF06
LyB06
Wol06
Liv06
Rwb06
DuB06
StB06
SCB06
RyB06
NeO06
CES06
InC06
Amb06
Eigenvalues
DAC3-14, 2006
d = 0.2
LyB07
Rwb07
InF07
Rye07
Amb07
Liv07
StB07
OffF07
Eigenvalues
DAC3-14, 2007
d = 0.2
Rye08
Ne Amb08
Liv08
InC08
StB08
Off08
LyB08
Inf08
Eigenvalues
DAC3-14, 2008


Eigenvalues
d = 0.2
NeD05
StB05
Amb05
LyB05
InC05
OffF05
DAC1-90, 2005
Eigenvalues
d = 0.2
WolI06
LIv06
DuB06
InC06
Amb06
Off06
RyB06
CES06
Rwb06
LyB06
StB06
InF06
DAC1-90, 2006
Eigenvalues
d = 0.2
LyB07
OffF07
Rye07
StB07
Amb07
NeD07
InF07
InC07
Rwb07
DAC1-90, 2007
Eigenvalues
d = 0.2
Rye08
StB08
LyB08
Liv08
Inf08
Off08
InC08
NeD08
DAC1-90, 2008

6.4.4 Multilocus pairwise population differentiation:

Since multilocus pairwise estimates of differentiation (UCEs and HTEs) in all years were very similar, both in pattern and values, only the results for 2008 samples are given as example (Table 6.4). In general, pairwise UCE values were close to zero and non significant between samples from the same sea basin (i.e. North Sea, Irish Sea), but higher (up to 0.019) and highly significant ($p<0.001$) when compared across sea basins for all years.

Standardization did not alter the ranking of pairwise differences, but increased the values of differentiation as a function of locus heterozygosity, which, as seen before, had variable effects across loci. Differentiation after correction remained negative or very small within-basin, while it increased by a factor of 3-7, and reached levels above 0.200 between North Sea and Irish Sea samples for several loci. Multilocus Θ_{WC} estimates (Rousset, 2008) and arithmetic average Θ_{WC} multilocus estimates were not noticeably different, and therefore only the latter are reported here. Pairwise G'_{ST_est}, D_{est} and φ'_{ST} produced very similar patterns of differentiation to those produced by Θ_{WC} and Θ'_{WC}, but in the same range of magnitudes as Θ'_{WC}. The effect of including individuals with missing data points was minimal for all estimators, while including loci *DAC1-35* and *DAC5-70* (which had suspected null alleles) modified the values and general pattern slightly. No general direction could be extracted from the effect of including the loci with null alleles in estimation of pairwise differentiation, which is what would be expected by randomly removing alleles from genotypes.

Table 6.4: Multilocus pairwise differentiation (Next page)

Pairwise estimators of differentiation between dab samples, sorted by estimators for the year 2008. Tables correspond to pairs of tradicional (above diagonal ↑) and heterozygosity corrected estimators (below diagonal ↓), except for Nei's and Jost' distances where ***Ds*** is above diagonal, and $\boldsymbol{D_{est}}$ below diagonal, and probabilities of random differentiation where **genic** are above diagonal and **genotypic** below diagonal. Significance values are denoted with stars (*=below 0.05; **=below 0.01; ***=below 0.001), while probabilities of random genic and genotypic Exact *G* test values below 0.05 are in **bold**. Differentiation values are colour coded from low (yellow) to high (orange) to facilitate interpretation.

***P Genic* (↑) vs. *P Genotypic* (↓)**

	NeD08	Amb08	Off08	Inf08	Rye08	LyB08	InC08	Liv08	StB08
NeD08	**0.001**	0.211	0.134	**0.002**	**0.008**	**0.016**	**<0.001**	**<0.001**	**<0.001**
Amb08	**0.006**		**0.038**	**0.004**	**0.019**	**0.002**	**<0.001**	**<0.001**	**<0.001**
Off08	0.293	0.213		**0.003**	**0.007**	**0.007**	**<0.001**	**<0.001**	**<0.001**
Inf08	**0.016**	0.103	0.058		0.117	**0.007**	**<0.001**	**<0.001**	**<0.001**
Rye08	**0.039**	0.126	0.065	0.363		0.096	**<0.001**	**<0.001**	**<0.001**
LyB08	0.066	**0.023**	**0.041**	0.112	0.301		0.194	**<0.001**	**<0.001**
InC08	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.002**	0.409		0.054	**0.001**
Liv08	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.474		0.080
StB08	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.035**	0.329	

Θ_{WC} (↑) vs. Θ'_{WC}

	NeD08	Amb08	Off08	Inf08	Rye08	LyB08	InC08	Liv08	StB08
NeD08		0.002	0.001 *	0.005 **	0.004 *	0.003	0.021 ***	0.011 ***	0.011 ***
Amb08	0.001		0.001	0.001	0.002 **	0.001	0.015 ***	0.006 ***	0.009 ***
Off08	0.009	0.001		0.001	0.000	0.004 *	0.017 ***	0.010 ***	0.007 ***
Inf08	0.013	0.002	0.004		0.000	0.004	0.017 ***	0.007 ***	0.007 ***
Rye08	0.010	0.005	0.003	-0.004		0.003	0.013 ***	0.009 ***	0.006 ***
LyB08	0.011	0.013	0.025	0.006	0.006		0.007	0.004 *	0.006 ***
InC08	0.047	0.053	0.042	0.049	0.030	0.023		0.001	0.003
Liv08	0.044	0.044	0.047	0.038	0.043	0.035	0.002		0.002
StB08	0.058	0.061	0.055	0.049	0.047	0.049	0.016	-0.002	

G_{ST_est} (↑) vs. G_{ST_est}' (↓)

	Ned08	Amb08	Off08	InF08	Rye08	LyB08	InC08	Liv08	StB08
Ned08		0.001	0.001	0.002	0.002	0.001	0.011	0.006	0.006
Amb08	0.000		0.001	0.000	0.001	0.001	0.008	0.003	0.005
Off08	0.009	0.000		0.000	0.000	0.002	0.008	0.005	0.004
InF08	0.011	0.002	0.004		0.000	0.002	0.008	0.004	0.004
Rye08	0.009	0.005	0.004	-0.003		0.001	0.006	0.005	0.003
LyB08	0.011	0.012	0.024	0.005	0.006		0.004	0.002	0.003
InC08	0.039	0.046	0.035	0.042	0.027	0.023		0.001	0.002
Liv08	0.041	0.042	0.044	0.037	0.041	0.036	0.007		0.001
StB08	0.054	0.057	0.052	0.046	0.045	0.048	0.017	0.000	

φ_{ST} (↑) vs. φ'_{ST} (↓)

	Ned08	Amb08	Off08	InF08	Rye08	LyB08	InC08	Liv08	StB08
Ned08		0.000	0.001	0.002 **	0.001 *	0.000	0.010 ***	0.004 ***	0.008 ***
Amb08	0.000		0.000	0.001	0.001 *	0.001	0.010 ***	0.004 ***	0.008 ***
Off08	0.003	0.001		0.000	0.000	0.002 *	0.009 ***	0.005 ***	0.007 ***
InF08	0.008	0.002	0.001		0.000	0.001	0.011 ***	0.005 ***	0.006 ***
Rye08	0.004	0.004	0.000	0.001		0.000	0.007 ***	0.005 ***	0.006 ***
LyB08	0.000	0.004	0.009	0.003	0.001		0.003	0.000 **	0.005 ***
InC08	0.034	0.037	0.031	0.037	0.025	0.010		-0.005	0.000
Liv08	0.016	0.015	0.020	0.017	0.019	0.001	-0.019		-0.002
StB08	0.028	0.030	0.023	0.023	0.021	0.018	-0.001	-0.006	

D_S (Nei_{72}) (↑) vs. D_{est} (↓)

	Ned08	Amb08	Off08	InF08	Rye08	LyB08	InC08	Liv08	StB08
Ned08		0.015	0.017	0.020 **	0.018 *	0.022	0.055 ***	0.033 ***	0.036 ***
Amb08	-0.001		0.014	0.015	0.017	0.024	0.055 ***	0.031 ***	0.037 ***
Off08	0.008	0.000		0.014	0.013	0.027 *	0.049 ***	0.033 ***	0.031 ***
InF08	0.009	0.001	0.003		0.014	0.022	0.053 ***	0.031 ***	0.030 ***
Rye08	0.007	0.004	0.004	-0.003		0.021	0.046 ***	0.033 ***	0.030 ***
LyB08	0.009	0.012	0.022	0.003	0.005		0.042	0.031 *	0.035 ***
InC08	0.029	0.039	0.027	0.035	0.021	0.019		0.027	0.029
Liv08	0.036	0.039	0.039	0.033	0.037	0.034	0.006		0.016
StB08	0.049	0.053	0.048	0.043	0.042	0.045	0.015	-0.001	

6.5 Discussion:

6.5.1 Single locus information content:

Correcting estimates of genetic differentiation for locus heterozygosity had an important impact on interpreting the patterns of locus-specific genetic structuring. Θ_{WC}, G_{ST_est}, and φ_{ST} values were all negatively correlated with H_E (average of UCEs = -0.244; p≈0.05), a pattern that became even more apparent when *DAC3-14* (which had a high UCEs despite high H_E) was removed (average r = -0.420; p=0.001). The slope of the relationship of UCEs and H_E was not very steep because differentiation among the samples was generally not strong: a feature typical of marine species (Hauser & Carvalho, 2008). As populations become increasingly differentiated across the genome, UCEs are more likely to pick up differentiation at low H_E loci than at highly diverse loci which will result in steeper slopes such as those found for sockeye salmon (r≈-0.708; p<0.001) (Olsen *et al.*, 2004), walleye Pollock (r≈-0.792; p=0.001) (O'Reilly *et al.*, 2004) and zebra mussels (r≈-0.900; p≈0.036) (Astanei *et al.*, 2005). These studies involved either anadromous species, which are known to be much more strongly structured than marine fish (Ward *et al.*, 1994; DeWoody & Avise, 2000; King *et al.*, 2001), or covered large parts of the species range, which maximises the chances of finding structure. At the opposite end of the scale, Larsson *et al.* (2007) reported higher differentiation (Θ_{WC}=0.002) with microsatellites than with allozymes (Θ_{WC}=0.001), contradicting the dependence of Θ_{WC} on heterozygosity. However, in systems with low levels of structuring, such as in herring, Θ_{WC} is more likely to detect at least minimal differences for microsatellites than for low diversity allozymes, since the former typically have higher information content, that is, are more genetically diverse.

The negative relationship between UCEs and H_E would suggest that, in the case of dab, low H_E markers contain more information than high H_E ones. However, HCE and D_{est} all had positive relationships with H_E (average r≈0.325; p≈0.009), more accurately representing the higher chances of allele-rich loci to show genetic structure among populations. When plotted against H_E all four HTE showed a similar pattern: low heterozygosity loci showing narrower levels of differentiation, while high heterozygosity markers exhibited a much broader range of differentiation levels, truly representing the increased probability of highly diverse markers to be structured among populations. The pattern is concordant with that observed in other studies (Carreras-Carbonell *et al.*, 2006), where highly variable markers, when corrected for heterozygosity, display higher levels of differentiation.

The linear relationship between Θ_{max} and H_E suggest that standardising Θ_{WC} for individual loci could be simplified as Equation 6.1:

$$\Theta'_{WC} = \frac{\Theta_{WC}}{1 - H_E}$$

Although such a correction might appear to inflate estimates of differentiation, it actually has very little effect on non-structured loci: *DAG2-90* remains negative and *DAC5-77* barely changes in value (Θ_{WC} and Θ'_{WC} < 0.002) despite high gene diversities (H_E= 0.9 and 0.5 respectively). A note of caution should, however, be exercised here: pairwise Θ'_{WC} calculations cannot be simplified using locus overall H_E, as each pairwise comparison has its own unique specific expected total heterozygosity (H_T) associated with it. Such a correction would, however, be fairly easy to include in existing packages.

There were more mismatches between the significant departures from random genic distribution and UCEs (i.e. Θ_{WC} values of 0.001 with p<0.001) than with HTE. The magnitude of correction (Δ_{HC}) coincides with the significance associated with the exact *G* test of genic and genotypic differentiation, where the markers with the most statistically significant differences in allele frequency among samples correspond to those with the greatest Δ_{HC}. Furthermore, corrected values are better predictors of locus performance in the CA: loci such as *DAG4-64* and *DAC1-90* correctly separated North Sea from Irish Sea samples for all years (despite Θ_{WC}~0.003), confirming their utility at detecting structure in dab and deserved increase in differentiation estimate (Θ'_{WC}~ 0.070 and 0.040 respectively).

The most surprising result was the near perfect match between the different HTE in locus-specific differentiation values and ranking according to information content. Nevertheless, the values provided by the different HTEs varied slightly. D_{est} seemed less effective at detecting differentiation across low heterozygosity loci, so that D_{est} values for *DAC5-5* and *DAC2-37* were consistently lower than estimates from HCE (Θ'_{WC}, G'_{ST_est}, and φ'_{ST}). If for example, the structure of locus *DAC5-5* in 2008 is examined, the number of alleles and heterozygosity are found to be low (N_A=6; H_E=0.186). The alleles are not randomly distributed across samples (p=0.001) which results in Θ_{WC} and Θ'_{WC} of around 0.020 and the CA also separates samples according to basin. However, D_{est} estimates the overall differentiation value at 0.007. The reduced sensitivity of D_{est} to detect differentiation at low levels of heterozygosity was also denoted by the steeper correlation between D_{est} and H_E (r=0.414; p=0.001) than HCE and H_E (r≈0.295; p≈0.021) (Figure 6.3; Figure 6.4). Small D_{est} values at the low heterozygosity end of the spectrum were also observed by Ryman & Leimar (2009), and

as explained by Jost (2009), are due to the very nature of the estimator: D_{est} measures differentiation based on how many alleles are different between populations, so perhaps at reduced number of alleles the power of detecting subtle structure is diminished compared to estimators more directly based on heterozygosity.

Although O'Reilly *et al.* (2004) highlighted the negative relationship between H_E and differentiation, it was assumed that because the estimated Θ_{WC} did not reach the maximum value (Θ_{max}), Θ_{WC} was not being constrained by low homozygosity. Essentially, Θ_{max} was viewed as the saturation limit, beyond which any further change in allele frequencies would be obscured by homoplasy and thus would be non-informative. However, a Θ_{WC} of one will only be achieved when two populations are fixed for one alternative allele each (Kalinowski, 2002a). From unity, Θ_{max} decreases with increasing number of alleles, and to reach a Θ_{max} of any value ($\Theta_{max} = 1\text{-}H_E$) two populations must not share any alleles, which corresponds to the intuitive definition of full differentiation (100% differentiated), and which can occur at any value between 0 and 1, depending on heterozygosity. Hence, the space between zero and Θ_{max} is the full range of values for Θ_{WC}, and therefore the standardisation of Θ_{WC} (Θ'_{WC}) results in a more appropriate indicator of the level of population connectivity, which is usually the target of molecular ecology studies.

Unfortunately, O'Reilly *et al.* (2004) did not report a locus-specific exact *G*-test of genic or genotypic differentiation that would portray more accurately how informative the loci used were. Given the elevated number of significant comparisons in the pairwise exact *G*-test (42 out 45 comparisons with a $p<0.001$), however, it is likely that much information was missed in the seven loci with reported H_E above 0.90. Indeed if one recalculates locus-specific Θ'_{WC} from their published Θ_{WC} and H_E using Equation 6.1, some of their high polymorphism loci which showed slight differentiation (Θ_{WC}~0.002) reveal Θ'_{WC} as high as ~0.060. Furthermore, many of these loci (namely *Tch5, Tch6, Tch8,* and *Tch14*) with increased Θ'_{WC} coincide with those that produced more significant pairwise comparisons in the exact *G*- test of population differentiation.

6.5.2 Population genetic differentiation:

Dab were not expected to display strong levels of differentiation at the scale covered in this study, but the stability of differentiation between sea basins among the temporal replicate samples, regardless of which estimator is used, conveys strong evidence of some degree of genetic and potentially, demographic independence between Irish Sea and North Sea samples.

Pairwise comparisons of HTE did not alter the multilocus differentiation values among samples within the same basin noticeably, corroborating the rather homogeneous genetic composition of dab within basins. On the other hand, multilocus differentiation estimates among samples from different sea basins increased consistently, suggesting that previously undetected differences at high heterozygosity markers exerted an impact. The values obtained for multilocus Θ'_{WC} were very similar to those obtained from Θ_{WC} when only the six loci with H_E below 0.800 were used (data not shown), indicating that the low Θ_{WC} values of loci with high heterozygosity, but structured among samples, were acting to reduce the multilocus Θ_{WC} estimates unnecessarily. The concordance of values between the pairwise differences for three of the HTE (Θ'_{WC}, G'_{ST_est}, and D_{est}) is reassuring. Interestingly, the differentiation values of Θ'_{WC}, G'_{ST_est}, and D_{est}, were very similar to those obtained with Nei's original D_S formula (Nei, 1972), which confirms Jost's (2009) suggestion that $D \approx D_S$ when differentiation is not very strong. Multilocus φ'_{ST} estimates were about half of those from the other estimators, but still larger than uncorrected values.

Many studies have reported higher levels of structure using low polymorphism markers (such as allozymes), where microsatellites failed. De Innocentiis *et al.* (2001) reported higher levels of structure in dusky grouper, *Epinephelus marginatus,* using nine allozymes (Θ_{WC}=0.214) than with seven microsatellites (Θ_{WC}=0.018). However, when one of the allozymes showing extreme structuring (ADA, Θ_{WC}=0.717) is removed and HCE calculated from their published data, microsatellites show higher levels of overall differentiation (Θ'_{WC}=0.087) than allozymes (Θ'_{WC}=0.064). Likewise, the unexplained low global Θ_{WC} values (~0.016) obtained in a study of poorly dispersive whelks, *Buccinum undatum*, with a negative relationship between heterozygosity and differentiation (r=-0.799; p=0.105) (Weetman *et al.*, 2006), fitted expectations better with a recalculated average global Θ'_{WC} of 0.069. Moreover a positive relationship between Θ'_{WC} with gene diversity emerges (r=0.636; p=0.249). On a study of zebra mussels, *Dreissena polymorpha*, five microsatellites with very high H_E (0.790 to 0.890) failed to reveal any population structure within the British Isles (Astanei *et al.*, 2005). Nevertheless, a later study employing eight moderately heterozygous allozymes on the same mussel samples revealed much higher substructuring (Gosling *et al.*, 2008), thus reanalysis of the microsatellite data correcting for heterozygosity might reveal previously undetected structuring.

In the low differentiation system studied here, the heterozygosity correction only changed the values slightly, but the limits imposed by high heterozygosity on Θ_{WC} become increasingly relevant as samples are more differentiated. The differentiation levels found by Carreras-Carbonell *et al.* (2006) between two subspecies of *Tripterygion delaisi* were apparently low (Θ_{WC} =0.066). Nevertheless, when heterozygosity was accounted for, differentiation increased over an order of magnitude (G'_{ST}=0.740), more concordant with expectations for divergence of subspecies.

In addition to the intrinsic relationship between traditional estimators and heterozygosity, Kalinowski (2002a) reported that the effective population size (N_e) had a significant effect on Θ_{WC}, where small populations reached higher levels of Θ_{WC} than large populations. The observation was explained by the propensity for large populations to harbour more alleles than small ones. Thus the latter have lower heterozygosities, which allow larger Θ_{max}, yielding higher differentiation as measured with traditional estimators for smaller populations than for larger ones. Such pattern was also acknowledged by Ryman & Leimar (2009) between equilibrium G_{ST} and within-sample heterozygosity (H_S). Although, such processes are tangentially linked with the buffering inertia of large populations to genetic drift, it adds another dimension to the explanation of small differentiation values found in many marine organisms, exemplifying the importance of correcting for heterozygosity when studying genetic structure of large populations with highly diverse microsatellites.

6.5.3 Mutation and information content:

Several studies have identified the problem of estimating F_{ST} from highly polymorphic microsatellites (Olsen *et al.*, 2004; O'Reilly *et al.*, 2004; Astanei *et al.*, 2005), but all assumed that the error was in the markers, rather than the estimator: by equating H_S to mutation bias, they concluded that mutation was responsible for the lack of differentiation among populations. Indeed, mutation, as the ultimate source of all genetic variability, plays an important role in determining the information content of loci. Microsatellites have dominated population genetics recently due to their typically high and largely neutral rate of mutations. However, such high mutability can compromise their effectiveness as proxies of demographic processes: hypothetically, mutation rates could reach levels where homoplasy (mutation to an already existing allelic state) would mask the allelic patterns imprinted by evolutionary processes (Estoup *et al.*, 2002a).

Whether the low UCEs values obtained with high diversity loci is due to problems in the estimators or homoplasy needs careful examination. The value of differentiation at a neutral

locus between two populations is determined by the balance between two opposing forces: on one side, drift will change the allele frequencies randomly, while on the other, migration between the populations will homogenise allele frequencies. A third force, mutation, plays a pivotal role in our capacity to detect the balance. If a mutation results in a new allele, then the increased number of variants for drift to act upon will increase the power of differentiation detection at a locus. On the other hand, if an allele reverts back into a pre-existing allele (homoplasy), then the mutation will mimic the effect of a migrant, producing the false impression that the two populations are still exchanging genes. Thus high mutation rates will be associated both with higher power to detect differentiation and an increased risk of assuming that shared alleles between populations are identical by descent. Although the level of homoplasy at a locus has indeed been found to be correlated with mutation rate (Estoup *et al.*, 2002a; Brandström & Ellegren, 2009), loci with a higher number of alleles may be more resistant to the effects of homoplasy as homoplasious events will be diluted among the large number of alleles. Indeed, Estoup *et al.* (2002a) concluded that high diversity loci were still more informative, regardless of homoplasy, than less diverse loci in exact tests of differentiation, and that the most marked effects of homoplasy would be detected in highly mutable loci with a tightly restricted allele states.

Explaining the low UCE of high diversity loci by homoplasy would imply that homoplasy is equally rampant and purposeful in *all* high diversity loci to generate a consistent lack of additional differentiation between independent populations at these loci. On the contrary, in the present study, several highly diverse microsatellites had very low UCE values, but showed concordant structure of samples across locus-specific multivariate analysis, and the allelic distributions were not random across samples (exact *G* test $p<0.001$) indicating differentiation among samples and suggesting that homoplasy is not always the only force associated with low UCEs observed in highly diverse microsatellites.

Furthermore equating heterozygosity and mutation rate may be misleading, as although two equally diverse loci may have had the same number of mutations resulting in the current polymorphisms (*estimated* mutation rate), the speed and time in which the loci have reached the current state may be very different (*real* mutation rate), and thus, the two markers could carry very different information about demographic history depending on when in evolutionary time the loci became polymorphic. For example, despite the heterozygosity correction, non-structured markers remained uninformative. Given that most loci show some degree of differentiation across the sampled range, it becomes interesting to know why a few

loci are so consistently non-differentiated. All microsatellites were tested in eleven other European flatfish species for cross-species amplification (Tysklind *et al.*, 2009b). All non-informative loci (*DAC2-28*, *DAG2-90*, and *DAC5-77*) amplified in four to eight other closely related species, suggesting the loci are older than the separation of the different species and that some mechanism (either homoplasy within a limited allele size range or balancing selection) is preventing the accumulation of demographic information in these loci. Conversely, highly informative loci for population structure in dab (*DAC3-14*, *DAC4-64*, *DAC1-90*, *DAC2-15*, and *DAC1-55*) failed to amplify in any of the other surveyed species, indicating these loci emerged and became polymorphic exclusively within dab's evolutionary history, and thus, carry information solely about demographic processes of dab.

6.5.4 Drift and information content:

As with any contentious statistical novelty, HTEs have been rejected by some authors (Ryman & Leimar, 2008; 2009) because G'_{ST} implies that different loci experience varying levels of drift, a process which is also observed in the current study. Their simulations show that loci with the same mutation rate reach the same G_{ST} value when mutation equilibrium is reached *regardless* of initial heterozygosity (Figure 4 in Ryman & Leimar, 2009). However, two issues emerge. First, the maximum heterozygosity considered was 0.450, whereas higher heterozygosities where G_{ST} is expected to behave very poorly, are not reported. And secondly, populations will have loci with different mutation rates, thus graphs A and C of their Figure 4 will occur simultaneously within the same population, and thus resulting in different G_{ST} values among loci in the same populations *even* at mutation equilibrium.

In their simulations of change of differentiation over time at different mutation rates, Ryman & Leimar (2009) (Figure 1), again reject G'_{ST} because the different mutation rates reach maximum differentiation at different times, and are suggestive of different amounts of drift. However, drift will have a much stronger effect on highly variable loci: those with high mutation rates, as defined to Ryman & Leimar (2009), will display such characteristics as lower allele frequencies and are thus more prone to relative change and extinction than alleles at low diversity loci. Indeed, in the same simulation, markers with high mutation rates (Hs=0.800) reach their maximum differentiation value (G_{ST} = 0.018 or G'_{ST} = 1) at which the populations no longer share any alleles after ca. 10^{-3} generations. Lower heterozygosity markers (Hs=0.286; Hs=10^{-3} and Hs=10^{-5}), conversely take considerably longer (ca. 10^{-4} generations) to reach a state of maximum differentiation. The speed at which differentiation occurs in the Figure 1 system for Ryman & Leimar (2009) is, however, exaggerated, as no

migration is assumed between populations, and both drift and mutation (which under the IAM used will only create new alleles) will work synergistically to increase differentiation. Thus in natural systems where the effects of one process could mask the other, the speed of differentiation is expected to be slower, albeit still conforming to the same principles.

Finally, as pointed out by Jost (2009), G_{ST} and other UCEs may still influence the estimation of migrant and past demographic events, but are completely unreliable at depicting differentiation between populations because G_{ST}=0.001 may represent a range of structuring from no differentiation to complete segregation of allelic states. Moreover, differentiation among samples should be descriptive of the allele diversity observed, and independent of any assumptions or equilibrium states.

6.5.5 Detection of loci under selection:

The implications of the limitations imposed by heterozygosity on F_{ST} estimates are far from trivial. Different areas of the genome are expected to display varying levels of estimated differentiation between independent populations (Nosil *et al.*, 2009), however, accepting Θ_{WC} as an estimator of differentiation implies that different regions of the genome have different values for *maximum* levels of differentiation depending on allele diversity (i.e. two hypothetical populations with no similarities in their genome could have estimated F_{ST} of any value between 0 and 1 despite their genomes being completely different) which immediately negates the comparison or averaging Θ_{WC} among loci without prior standardisation.

Following the same line of thought, the description of markers under differential and balancing selection by comparing F_{ST} to heterozygosity would need re-defining. Under current assumptions an average Θ_{WC} is used to generate a 95% confidence interval *envelope* of possible Θ_{WC} across a range of H_E. The average Θ_{WC} value has been fixed when computing the simulated markers to remain the same, regardless of H_E, in an attempt to mimic real genomic differentiation (F_{ST}). However, as populations become differentiated, loci with lower H_E will generate higher Θ_{WC} values than high heterozygosity loci. Such a pattern can be observed in the *envelope* of Θ_{WC} for simulated loci, as it gets narrower at the high heterozygosity end of the spectrum. However, the key issue here is that the same value of Θ_{WC} implies very different levels of differentiation across the range of heterozygosity: a Θ_{WC} = 0.05 at H_E = 0.4 corresponds to many more alleles in common (8% of maximum differentiation) than Θ_{WC}=0.05 at H_E=0.9 (50% of maximum differentiation).

HTE would exhibit a different behaviour: at the same level of true genomic differentiation HTE are expected to increase (as would the height of the *envelope*) with H_E, given that there are more alleles for drift to act upon. In the case of dab, one of the most structured loci, *DAC3-14*, showed marked allele frequency differences in samples from different basins, and LOSITAN highlighted the locus as potentially under selection, a pattern also obvious when Θ_{WC} was plotted against H_E. However, once corrected for heterozygosity, estimates of differentiation in other loci resembled those in *DAC3-14*. Under this scenario, *DAC3-14* would most likely be within the 95% confidence *envelope*. Furthermore, if selection was effectively driving the structure found at *DAC3-14*, the expected direction of differentiation would be different to that shown by the rest of the markers (Schlötterer, 2002). The correspondence analysis revealed that several loci showed structure in the same direction as *DAC3-14* (i.e. separating Irish Sea from North Sea), suggesting a genome-wide differentiation rather than locus-specific selection.

Another example of the effect of heterozygosity correction on the evaluation of loci under selection can be observed in the Olsen *et al.*, (2004) study. Outstanding differentiation for both the least diverse single allozyme and microsatellite (locus-specific Θ_{WC} around 0.150 compared to the average of 0.023) and a strong and significant inverse correlation between Θ_{WC} and H_S (r=-0.805; p<0.001) were reported. However, when HCE are calculated using Equation 6.1, the average Θ'_{WC} and G'_{ST_est} are approximately 0.120 and the two previously outlying markers are no longer discordant with the rest. The negative correlation between differentiation and heterozygosity is also lost (r=0.217; p=0.344).

6.5.6 Comparison between estimators:

The function of an estimator is to find a value that best represents the relative time of divergence and the amount of gene flow between two populations (F_{ST}). It, thus, becomes essential that estimators increase linearly with time or reduced gene flow. A good estimator of differentiation should ideally be able to detect differentiation equally across the whole range of heterozygosity values (0-1), which would result on a horizontal line when averaging across loci with different heterozygosities. However, as alleles in more diverse loci have lower frequencies, they have a higher chance of being affected by drift, which results in diverse loci showing increased divergence, and a skewed slope for the across-heterozygosity loci average.

Traditional estimators fail to convey information about the organisation of alleles at high diversity markers, thus should not be deemed as appropriate estimators of real differentiation (F_{ST}) when using highly diverse microsatellites. Corrected versions of the

traditional estimators and new measures fare better at the high heterozygosity end of the spectrum. Nevertheless they differ subtly.

All four HTE gave very similar values for most markers, thought, G'_{ST_est} and D_{est} seemed more affected by null alleles in those loci surveyed than other estimators. Furthermore D_{est}, seemed less sensitive to structure in low diversity loci than its counterparts. Although φ'_{ST} locus-specific values were very similar to those computed with the rest of estimators, pairwise comparisons resulted in reduced values compared to other HTEs.

Keeping Θ_{WC} as the foundation for an HTE may offer some advantages. The popularity of Θ_{WC} is based on several studies that have demonstrated its reliability: Θ_{WC} has been found to have reduced bias and variance (Weir & Cockerham, 1984; Raufaste & Bonhomme, 2000), and was found to be relatively robust to differences in mutation model (Balloux & Lugon-Moulin, 2002) and mutation rate (Kalinowski, 2002a). Nonetheless, it is possible that the reduced variance and standard error of Θ_{WC} for loci with higher number of alleles (compared with D_S, D_A and D_C) reported by Kalinoswki (2002) are due to the reduced space available for variance (zero to Θ_{max}). Such effects mean that Θ'_{WC} will not display such advantages, and behave equally variably for high polymorphism loci as the other estimators. The only way to address such issues effectively is to evaluate the behaviour (variance and bias) of the estimates with simulations under different scenarios including different sample sizes, mutation models and rates, population models, and levels of homoplasy and null alleles.

6.6 Conclusion:

Traditional estimators of differentiation were compared empirically with high-heterozygosity-tolerant measures of differentiation on a dab data set. Results suggest that Θ_{WC}, G_{ST_est} and φ_{ST} consistently yielded reduced values for high heterozygosity markers, confirming their compromised performance in highly genetically diverse systems. Correcting for heterozygosity revealed that highly polymorphic markers contained considerable levels of undetected diversity, and could indeed be more informative regarding population structure than their less diverse counterparts. It is thereby suggested that heterozygosity-tolerant estimators be employed more widely when the aim is to assess the magnitude of differentiation, and particularly when highly variable microsatellites in marine species are analysed. However, their use for more complex analysis, such as estimation of number of migrants, population effective size or demographic events may need further careful consideration (Ryman & Leimar, 2008; 2009; Jost, 2009).

Of the measures considered here, Θ'_{WC} and φ'_{ST} were found to be the most reliable across the whole heterozygosity range and in loci with null alleles; however, simulations of the behaviour of corrected estimates under diverse scenarios (mutation model and rate, homoplasy, null alleles, and populations models) are required to assess their accuracy and stability.

Chapter 7: Population genetic structure of dab (*Limanda limanda* L.): a key biomonitoring species

7.1 Abstract:

The use of bioindicator fish species in monitoring programmes is widely accepted as a means of assessing marine ecosystem health. The effects of pollutants and other anthropogenic impacts on the health of individual fish is evaluated and compared to that of non-exposed reference fish. In the UK, dab, *Limanda limanda*, have been routinely used as environmental bioindicators of pollutant exposure, but little information exists on the population structuring or lack of between the sampling locations. In the current study, the genetic structure of dab and its temporal stability around the British Isles is evaluated with 16 microsatellite loci over the course of four sampling years. Several analytical methods coincide in defining two significant and clear groups corresponding to the North and Irish Seas. Further structuring within either sea basin and in the English Channel was weaker and fluctuated over time. Two independent methods to evaluate past demographic changes concluded that a significant population expansion signal could be detected in most locations. The consequences of comparing genetically structured populations in their environmental response are briefly discussed.

7.2 Introduction:

Awareness that the natural environment is not just a precious commodity, but intrinsic to the sustained existence of human societies and the livelihoods of its members, is increasingly commonplace (Islam & Tanaka, 2004; Hughes *et al.*, 2005; Schiedek *et al.*, 2007; Kite-Powell *et al.*, 2008; Stott, 2009). Human activities often result in the transformation or contamination of habitats with high levels of pollutants, which may have deleterious effects on wildlife, ecosystems, and ultimately on human uses of such ecosystems (Tanabe, 1988; Kalantzi *et al.*, 2001). Aquatic environments are particularly threatened because they often serve as intentional, or more often, unintentional repositories of pollutants (Haynes & Johnson, 2000; Islam & Tanaka, 2004). Such anthropogenic effluents later concentrate in lakes, estuaries, coastal waters and ultimately oceans, where levels may reach more significant levels (Ballschmiter, 1992; Scheringer, 2009). Heavily polluted environments can have adverse effects on human activities, and, within the marine environment, the most

conspicuous impacts include poisoning of food items (fish and shellfish) (Judd *et al.*, 2004; Corsolini *et al.*, 2005), reduced fish yields and loss of leisure areas (Jones, 2006), and loss of biodiversity and ecosystem services (Worm *et al.*, 2006).

Governments are now urged to evaluate levels of pollution in aquatic environments, which in turn has resulted in the development and implementation of suites of tools and protocols for detection and remediation of anthropogenic impacts. The UK aims for "clean, healthy, safe, productive and biologically diverse ocean and seas" (DEFRA, 2002) has resulted in the creation of the Marine and Coastal Access Bill (DEFRA, 2009). At a wider European level, the OSPAR convention (OSPAR Commission, 2000) and the European Union Marine Strategy Framework Directive (European Parliament, 2008) promote coordination of environment management between member states.

One way of evaluating pollution levels, is to employ bioindicator species for which parameters such as presence or absence, population size, health status, and other proxies of ecosystem health are regularly recorded providing time series data (Phillips & Segar, 1986; Whitfield & Elliot, 2002; van der Oost *et al.*, 2003; Breine *et al.*, 2007). Changes in such parameters can be analysed in conjunction with changes in pollutant levels and other environmental variables, to ascertain the effect of pollutants on natural ecosystems, their trends over time, and the efficiency of endorsed environmental policies (Vethaak *et al.*, 2009). Standard techniques on how to collect and analyse biomarker data (Feist *et al.*, 2004) and accompanying quality assurance procedures (BEQUALM, 2009), have been compiled at both European and UK levels into biomonitoring programmes such as the Joint Assessment and Monitoring Program, JAMP (OSPAR, 2009), and the Clean Seas Environmental Monitoring Programme, CSEMP (CEFAS, 2003c).

Due to their benthic lifestyle, and thus, close proximity to settled pollutants, flatfishes are considered prime candidates for bio-assessing pollution in estuaries and coastal waters (MAFF, 1987; Köhler *et al.*, 1992; Förlin & Celander, 1993; Reichert *et al.*, 1998; Stentiford *et al.*, 2003; Feist *et al.*, 2004; Leonardi *et al.*, 2009). Furthermore, their medium-high trophic status, which renders them prone to bioaccumulation of pollutants (Hellou *et al.*, 1994; Hellou & Warren, 1997; Nakata *et al.*, 2003), and relative longevity (Deniel, 1990), which permits the development of long term diseases (Feist *et al.*, 2004), also enhances their attractiveness in biomonitoring schemes. Thus, in UK shallow waters, dab (*Limanda limanda*) is used as a key bioindicator sentinel species (Feist *et al.*, 2004). Detailed biomonitoring of dab is now performed on a regular basis, and has excelled as a way to assess levels of

pollutants and its effects on individual fish (CEFAS, 2003c; Stentiford *et al.*, 2003; Skouras *et al.*, 2003; Feist & Stentiford, 2005; Vethaak *et al.*, 2009; Stentiford *et al.*, 2009). Nonetheless, the consequences of pollution at the population level within an evolutionary timescale on dab have received relatively little attention. Different populations may exhibit differing susceptibility thresholds, and in the long term, the selection pressure imposed by pollutants can have important consequences on individual responses to pollutants and the genetic composition of populations (Theodorakis & Shugart, 1997; Belfiore & Anderson, 2001; Bickham *et al.*, 2000; Gardeström *et al.*, 2006; Nowak *et al.*, 2009). In addition to affecting valuable population genetic resources (FAO, 2008), and thus, resilience (Reusch *et al.*, 2005), varying susceptibility and adaptive changes in the pollution tolerance of wild fish could potentially impinge on the interpretation of biomonitoring data (Chapter 1).

Furthermore, important questions about dab biology remain unanswered, such as patterns of population connectivity and the stability of such population structure over time. These concepts are of paramount importance to evaluate the potential for adaptation of populations to local conditions (Carvalho, 1993; Conover *et al.*, 2006). In the same way as different individuals respond to pollutants in different ways, so do populations, thus it is important to estimate the extent and boundaries of biologically meaningful groups of interbreeding individuals. In doing so, the response of fish to contaminants can be controlled for varying levels of dispersal, susceptibility, or localised selection factors (either natural or anthropogenic). Indeed, the extent to which individuals share a common gene pool, as well as factors such as population size and the intensity of selection pressures such as pollution, will not only determine the nature and speed of response to pollution events, but also importantly the resilience, or ability of populations to recover from localised declines (Bickham *et al.*, 2000; Reusch *et al.*, 2005). Therefore, information on population structure, genetic variability, connectivity, temporal stability and adult or larval migration between locations are all needed for more meaningful interpretation of biomonitoring data (Theodorakis, 2001).

Detecting structure by genetic means in marine organisms is not an easy task. The resistance of large sized populations to genetic drift, which minimises the process of genetic differentiation, together with high fecundity and the potential for extended dispersal, nurture the perfect conditions for hypothetical panmixia (Ward *et al.*, 1994; Waples, 1998). Despite such potential, the elusive structured nature or many marine fish has been revealed with the use of genetic markers, particularly microsatellites (Carvalho & Hauser, 1994a; Lundy *et al.*,

1999; Ruzzante *et al.*, 1999; Bernal-Ramírez *et al.*, 2003; Nielsen *et al.*, 2004; Jørgensen *et al.*, 2005; Mariani *et al.*, 2005; Hemmer-Hansen *et al.*, 2007b; Limborg *et al.*, 2009).

Here, the genetic diversity of dab using a panel of species-specific microsatellites (Tysklind *et al.*, 2009b) will be examined and the null hypothesis that dab are genetically homogeneous in the coastal waters around the British Isles will be testes. Alternatively, the distribution of genetic diversity of the species across the geographical range covered by the CSEMP programme will be dissected, an approach commonly termed *landscape genetics* (Manel *et al.*, 2003) and successfully applied to marine fish species (Jørgensen *et al.*, 2005; Hansen & Hemmer-Hansen, 2007; McCairns & Bernatchez, 2008; Selkoe *et al.*, 2008; Galarza *et al.*, 2009). The genetic signatures of past demographic events in relation to the extant distribution of genetic diversity will also be explored. Finally, the temporal stability of patterns of detectable genetic structuring will be assessed, which helps to interpret the biological significance of patterns of genetic differentiation (Carvalho & Hauser, 1998; Waples, 1998; Bernal-Ramírez *et al.*, 2003; Mariani *et al.*, 2005; Nielsen *et al.*, 2009).

7.3 Materials and Methods:

7.3.1 Sampling:

Since the mid 1980's dab have been monitored every year around the UK for the presence of parasites, diseases, and other bioindicators of pollutant exposure (MAFF, 1987; 1995; CEFAS, 2000; 2003a; CEFAS, 2005; Feist *et al.*, 2008). Currently the CSEMP operates for three weeks at the beginning of each summer and has several established sampling stations which are visited during that period (CEFAS, 2005). Standard numbers of fish are processed for different projects, and each individual fish is allocated an identification number so that data from different projects can be cross-referenced. Fin clips were collected from these fish and stored in absolute ethanol for genetic analysis. Genetic samples were collected for four consecutive years (2005-2008) in up to 15 stations covering four areas: North Sea (ICES areas: IVb, IVc), English Channel (ICES areas: VIId, VIIe), and Irish Sea (ICES area: VIIa). The latter includes samples from both Cardigan Bay, hereafter as Irish Sea (South), and samples north of Anglesey, hereafter Irish Sea (North). Complementary to the biomonitoring programme samples, three additional samples were obtained from the South Celtic Sea (ICES area: VIIj) and the Atlantic coast of Ireland (ICES areas: VIa, VIIb). Overall, 40 samples, each with 21 to 183 individual dab (and a grand total of 3027), were genotyped (Figure 7.1; Table 6.1). Samples were allocated a three letter code representing the sampling location and two

numbers representing the year of collection (05-08) (Table 6.1). DNA was extracted from the fin clips using the hi-salt extraction method (Aljanabi & Martinez, 1997), and samples genotyped for 16 loci published in Tysklind *et al.* (2009b) in multiplex PCRs as described in Chapter 4. Samples collected in 2005, and some of those collected in 2006 were DNA extracted externally by qualified CEFAS technicians.

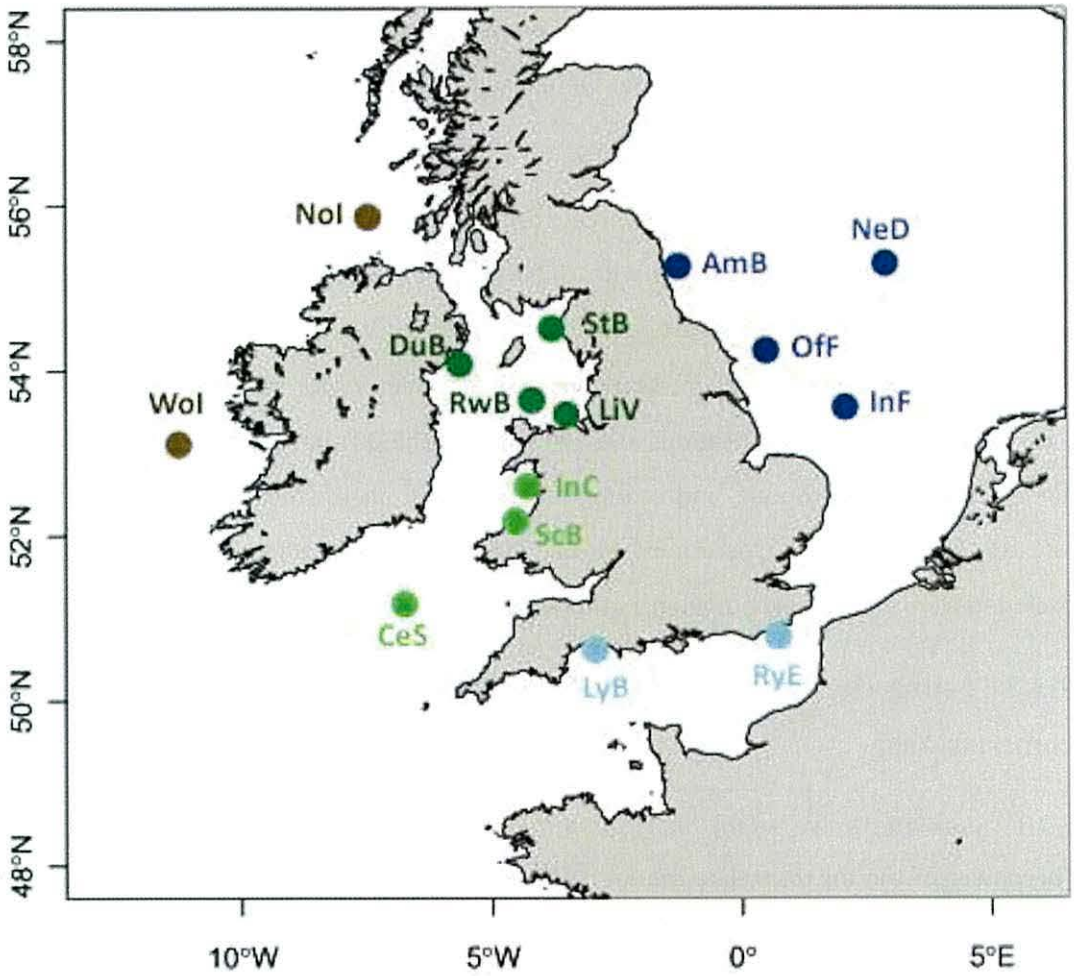


Figure 7.1: Sampling locations.
Dab sampling locations around the British Isles colour coded by basin. North Sea sites are in dark blue (North east Dogger Bank: NeD; Amble: AmB; Off Flamborough: OfF; Indefatigable Bank: InF); English Channel sites are in light blue (Rye Bay: RyE; Lyme Bay: LyB); Irish Sea (South) sites are in light green (Celtic Sea: CeS; South Cardigan Bay: ScB; Inner Cardigan Bay: InC); Irish Sea (North) locations are in dark green (Red Wharf Bay: RwB; Liverpool Bay: LiV; Saint Bees Point: StB; Dundrum Bay: DuB); Atlantic samples are in brown (North of Ireland: NoI; West of Ireland: WoI).

7.3.2 Data quality assurance, locus characteristics and conformity to expectations:

Allele sizes were determined with GENE MAPPER® and each individual genotype reviewed at least twice. Each individual peak (over 100,000 for those individuals included in the analysis) was carefully checked for correct size standard labelling and allele miscoring or cross-colour

pull-ups (an artefact peak created by interference between spectral absorbances) (Rudin & Inman, 2002). Any rare or unexpected alleles or HWE and linkage disequilibrium outliers were thoroughly checked once again. One sample, ScB06, was DNA-extracted and genotyped twice, at the beginning and end of the project to check for scoring consistency (DeWoody *et al.*, 2006). At least two individuals per plate were re-genotyped to verify amplification consistency between plates. Some individuals could not be fully genotyped even after several attempted amplifications and have been removed for some of the analyses. Sample sizes after removal of incomplete genotypes were also calculated (Table 6.1).

The software CREATE was used to produce input files whenever possible (Coombs *et al.*, 2008). Observed and expected heterozygosity, and Locus-specific Θ_{WC} and f_{wc} (Weir & Cockerham, 1984) were estimated with GENALEX (Peakall & Smouse, 2001) and GENEPOP V4.0 (Rousset, 2008). The same two software packages were used to check genotypes for HWE and linkage disequilibrium. Markov Chain parameters for the HWE test in GENEPOP V4.0 were 10,000 dememorisations and 100 batches of 5,000 iterations. The presence of null alleles, large allele dropout, and scoring errors due to allele stuttering were assessed with MICRO-CHECKER (van Oosterhout *et al.*, 2004). Loci suspected with null alleles were excluded from analysis except where otherwise stated.

7.3.3 Testing the existence of differentiation among samples and pairwise compassions:

7.3.3.1 Simulated power of the microsatellite loci to detect differentiation:

The power of the microsatellite marker suite to detect differentiation between independent populations was assessed with POWSIM (Ryman & Palm, 2006). The default Markov chain parameters were used with the overall combined allele frequencies, dividing them into two populations with 10,000 individuals as effective population size (*Ne)*, and for varying number of generations (proportional to *Fst*) and sampling sizes (100 and 50 individuals per sample).

7.3.3.2 Testing for the existence of structure among samples:

Global Θ_{WC} was calculated with FSTAT 2.9.3 (Goudet, 1995; 2001). Pairwise Θ_{WC} (Weir & Cockerham, 1984) were calculated in GENEPOP V4.0. Associated significances based on 1000 permutations were calculated with GENETIX (Belkhir *et al.*, 1996-2004). Each year was treated independently as suggested by Balloux & Lugon-Moulin (2002). As seen in Chapter 6, traditional estimators of differentiation (Θ_{WC}) are heavily influenced by heterozygosity and lose their linearity with actual differentiation at high levels of within-sample heterozygosity

(Shriver *et al.*, 1995; Hedrick, 1999b; Hedrick, 2005). Therefore, and given that several loci exhibited high levels of allele diversity and associated heterozygosity, a standardised estimator of differentiation, Θ'_{WC}, was calculated. Standardised Θ'_{WC} estimates were calculated in a similar way as G'_{ST} (Hedrick, 2005), and as hinted by Meirmans (2006). Θ_{max} was calculated by transforming the raw genotype data with RECODEDATA V.0.1 (Meirmans, 2006), so that all populations had non-overlapping allele ranges for all loci. In these conditions, all populations are maximally differentiated (no alleles in common) but heterozygosities remain unaffected. Locus-specific Θ_{max} was calculated with GENEPOP V4.0 (Rousset, 2008). Standardized Θ'_{WC} was calculated by dividing the original Θ_{WC} by Θ_{max}. Other studies have also calculated standardised Θ'_{WC} estimates in a similar way (Kenchington *et al.*, 2009; Papetti *et al.*, 2009). Multilocus average (arithmetic mean of all loci) was calculated for Θ_{WC} and for Θ'_{WC} allowing a comparison between both estimators.

Exact *G*-test (Raymond & Rousset, 1995) of the probability of random allele and genotype distributions across samples (genic and genotypic subdivision) were calculated with GENEPOP V4.0 (Rousset, 2008), where the Markov Chain parameters were set to 10,000 dememorisations and 500 batches of 10,000 iterations.

The population genetics programme GENODIVE (Meirmans & van Tiernderen, 2004) offers a clustering algorithm which aims at grouping a number of observations (either population samples or individuals) into *K* most probable clusters based on observation means. Observations are then assigned to clusters with the closest mean. The analysis, named K-Means clustering, was employed to investigate the most likely number of genetic clusters within the data and which samples belong to each cluster. Analysis of molecular variance (AMOVA) estimating φ_{ST} (Excoffier *et al.*, 1992) and the heterozygosity corrected φ'_{ST} (Meirmans, 2006) were performed within the same software, where significance was assessed with 10,000 permutations. Sources of variation included within individuals, among individuals within samples, among samples within clusters (as estimated by the K-means clustering analysis), and among clusters. AMOVAs were run for each year separately, for all years together but subdivided by clusters, and with all data using year as clusters.

7.3.3.3 Genetic relationship between samples:

Neighbour-joining trees were constructed with the PHYLIP package (Felsenstein, 1989) based on Nei's Genetic Distance, D_S (Nei, 1972), which is nearly equivalent to the differentiation calculated with heterozygosity corrected estimators when differentiation is small (Jost, 2009).

Robustness of the nodes was assessed by bootstrapping the allelic frequencies at different loci 1,000 times.

7.3.4 Geo-referenced analysis:

7.3.4.1 Relationship between genetic distance and geographic distance:

Patterns of genetic differentiation can sometimes be explained by geographic distance between sampling locations (Kotoulas *et al.*, 1995; King *et al.*, 2001; Mariani *et al.*, 2005), a phenomenon called isolation by distance (Wright, 1943). Therefore, the influence on the genetic structure of the distance needed for a dab to migrate between areas will be evaluated. Patterns of isolation by distance were tested using Mantel tests in IBD V.1.52 (Bohonak, 2002). Regressions of *Fst*/(1-*Fst*) over the log of the minimum distance by sea (km measured in Google™ Earth) were tested with 100,000 randomisations as suggested by Rousset (1997). Both Θ_{WC} and Θ'_{WC} estimates were used for comparison. In order to unravel the effects of geographic distance versus sea basin on the genetic distance, partial correlation mantel tests with a third matrix containing sea basin (North Sea-Irish Sea-Atlantic), in which 0 coded for samples within the same basin and 1 for different basins, were performed. Mimicking McCairns & Bernatchez (2008), Mantel tests of correlation between genetic and geographic distance of samples within basins were computed to evaluate whether the isolation by distance held within basin. LyB samples were excluded from the latter analysis due to their undifferentiated nature (see results). The basin-wise test was not possible for 2005 due to the low number of samples available.

7.3.4.2 Evaluation of the most important barriers to gene exchange between locations:

Aiming to better understand the patterns of differentiation between sampling locations, the software BARRIER v2.2 (Manni & Guérard, 2004; Manni *et al.*, 2004) was employed. The programme generates cells around the sampling locations which have edges in common with other sampling locations, thus a network of connectivity between the geo-referenced locations is created. The sharpest changes in genetic composition between neighbouring locations can then be assessed. The interconnectivity network can be modified to depict realistic patterns of interaction between sampling locations (i.e. when investigating marine organisms, the "edges" between samples collected at either side of a mass of land are not allowed). The coordinates of the dab sampling location were plotted and two different networks investigated: one in which sampling locations in the Irish Sea connected to the North Sea, and one in which empty cells (virtual points) were used to represent Great Britain and Ireland, effectively eliminating the common edges between Irish Sea and North Sea, and

between OfF and RyE, and LiV and LyB. Edges between StB and RwB and InC, and between AmB and NeD were allowed. The different genetic distance matrices calculated in Chapter 6 (Θ_{WC}, Θ'_{WC}, G_{ST_est}, G'_{ST_est}, and D_{est}) were plotted onto the map and compared.

Additionally, the software can take several different matrices and provide a support value for each barrier based on how many matrices coincide in supporting each barrier. Therefore, two multi-matrix analyses were performed: first, an evaluation of the concordance between distance matrices of each locus, giving an estimate of how much of the genome support a certain barrier; and secondly, a significance value for each barrier can be obtained by analysing multilocus bootstrap replicates. Bootstrapping of corrected estimates of differentiation was unpractical, therefore Nei's D_S (Nei, 1972) was used instead as it is very similar to corrected estimates of differentiation in low differentiation systems (Jost, 2009). The significance of the barriers was calculated by resampling 1,000 bootstrapped matrices created with the PHYLIP package (Felsenstein, 1989).

7.3.4.3 Population membership of geo-referenced samples:

Each sample does not necessarily represent unique populations, as several samples can be collected from the same population. Therefore the membership of each sample to populations was studied in two different ways: first employing Bayesian algorithms (GENELAND) and secondly, to provide a contrasting outlook, with a principal component analysis (ADEGENET).

The R package GENELAND (Guillot *et al.*, 2005; 2005; 2008; Guillot, 2008) carries a Bayesian clustering algorithm to estimate the number of populations based on conformity to HWE and linkage equilibrium developed by Pritchard *et al.* (2000), and modified to include geo-referenced data. The outputs are presented as estimated number of populations (k) and Voronoi tessellation maps of posterior probability of belonging to each of the estimated populations. Each sampling year was analysed independently, with 500,000 iterations and thinning to every 100th, the correlated allele model, and up to 20 populations. As the GENELAND algorithm has been designed to cope with null alleles, all loci were included in the analysis. Males and females were also analysed separately to investigate the possibility of sex-biased dispersal. When making plots, the coordinates of LyB samples were modified slightly towards the south (Lat: 50.61 changed to Lat: 50.41) in order to improve the visualisation of the membership of the sample when the map of Britain was overlaid. The correction was applied only when plotting, so it does not have any effect on the calculation of membership.

Another two R packages, ADE4 (Chessel *et al.*, 2004) and ADEGENET (Jombart, 2008) were invoked to perform multivariate analysis at several levels of the data (Jombart *et al.*, 2009). Sample correspondence analysis (CA) of the whole data set (All years) and spatial principal component analysis (sPCA) for every year were studied to understand the clustering and positioning of the samples in respect to others with and without spatial priors (Jombart *et al.*, 2008). Plots were centred on the origin and missing genotypes were treated as suggested in the ADEGENET manual: replaced with the allele means for PCA and the mean χ^2 distance for CA, which effectively places missing data at the origin of the axis. Analyses were performed also using only fully genotyped individuals, but results were the same as those from the corrected data set (allele means or χ^2). For the sPCA, the networks between samples were modified as with BARRIER to represent biologically realistic connections (i.e. no direct links between Irish Sea and North Sea samples).

7.3.4.4 Migration and admixture:

To assess the levels of extant connectivity among basins, the probability of individuals being first generation migrants was assessed with GENECLASS2 (Piry *et al.*, 2004). Since not all possible dab populations were sampled, only the likelihood of an individual being born where it was caught was computed (L_h). The allele frequencies-based method (Paetkau *et al.*, 1995) with a default frequency for missing alleles of 0.001 was used as criteria for computation of L_h values. Rannala and Mountain's (1997) multilocus Bayesian method was also computed for comparison. The associated probability of L_h values was computed by simulating 10,000 individuals (Paetkau *et al.*, 2004). The analysis was run separately for each year with only individuals fully genotyped.

Individuals from one of the samples (OfF05), which was received as DNA extracts, were located in between individuals from StB05 (i.e. individuals 1 to 12 were labelled as StB05, while individuals 13-42 were labelled as OfF05, and individuals 44-95 were again labelled as StB05). Due to the low amplification success of both the StB05 (18) and OfF05 (16) some extra samples were requested and five individuals were sent. These latter individuals were labelled as OfF05x thereafter. Since then the OfF05 sample was treated with caution. To evaluate the provenance of OfF05 individuals, a GENECLASS2 (Piry *et al.*, 2004) assignment exercise was performed. Reference population allele frequencies were created from the individuals fully genotyped in 2005 from all other samples. The individuals from OfF05 were then compared to the rest of the populations to check whether they were more likely to come from the North Sea or the Irish Sea.

The admixture proportions at the population level, estimated as m_Y (Bertorelle & Excoffier, 1998), of the LyB and InC samples were evaluated in ADMIX 2.0 (Dupanloup & Bertorelle, 2001), using two grouped reference samples: North Sea (NeD and AmB) and Irish Sea (LiV and StB). Each year was analysed separately except for 2005, due to small sample size. Instead, the Irish Sea samples of 2006 were used as Irish Sea reference for 2005, while the North Sea samples of 2005 were kept as North Sea reference for the 2005 admixture analysis. Standard deviations were estimated with 10,000 bootstraps over loci. Since only very recent admixture events were targeted, molecular distances (measured as allele size difference under the SMM) are unlikely to play an influential role, and thus ignored in the current analysis (Hansen *et al.*, 2008). The admixture proportions of two samples (InF07 and RwB07) considered to be not influenced by migrants were also calculated as a control.

Maximum-likelihood estimates of individual admixture proportions (h) were estimated with the Buerkle (2005) method in GENODIVE (Meirmans & van Tiernderen, 2004). All of the NeD and AmB, and LiV and StB samples (except StB05) were pooled to produce reference (North Sea) and alternative (Irish Sea) populations respectively.

7.3.5 Temporal analysis:

The existence of significant genetic structure does not necessarily imply that it is biologically meaningful (Waples, 1998), thus, the temporal stability of the genetic structure is key in discarding finding structure by chance distribution of alleles (Carvalho & Hauser, 1998). Several methods were employed to assess the stability of genetic structure over the temporal replicate samples. First, pairwise Θ_{WC} and Θ'_{WC} were calculated for samples within location across years, and correlation between Θ_{WC} and time (years) were studied with Mantel tests in IBD V.1.52 (Bohonak, 2002). To evaluate if genetic distance between samples were stable across years, correlation between years of genetic distance (Θ_{WC}) across the whole study area was also tested with Mantel tests.

In order to generate a graphical representation of temporal structure and connectivity within and among locations, an artificial geographic and temporal matrix was created. The GENELAND manual's suggestion of linearizing 2-dimensional coordinates into 1-dimensional distances (Guillot, 2008) was exploited to transform geographical coordinates to one dimension. The distances of sampling locations to an arbitrary central point (LyB) was used as the abscissa and sampling year as the ordinate. The geographical distance was divided by ten so abscissa and ordinate were of similar magnitudes. The GENELAND algorithm was run for four independent Markov chain Monte Carlo (MCMC) with 1,000,000 iterations sampled

every 200th and discarding the first 500. The analysis was repeated three times with a random order of sampling locations. The same grid was analysed with a sPCA in ADEGENET, in which the connection network was modified to allow all geographically possible connections across all years. Care was taken to ensure that links between samples from the same location, but from different years, were allowed.

7.3.6 Age and sex effects on genetic differentiation:

Age data, read from otoliths, were supplied by CEFAS for three samples in 2007 (NeD07, RyE07, LiV07) with 178 to 184 individuals each. To evaluate whether there were cohort changes in genetic composition within locations, the samples were subdivided into age classes and an analysis of genetic differentiation among age classes within location was performed for all subclasses with at least 20 individuals. To assess whether sex-biased dispersal was prevalent in dab, all individuals were visually sexed upon collection, and genotypes were subdivided accordingly. Genetic differentiation between males and females within and among sampling sites was investigated using neighbour-joining trees (Felsenstein, 1989), and GENELAND plots (Guillot *et al.*, 2005).

7.3.7 Past demographic events:

Changes in population size remain imprinted as signatures on their genetic diversity. In a constant sized population the allelic diversity emerges randomly throughout time. On the other hand, a population experiencing rapid increase in the number of individuals (effectively a population expansion) will undergo an increase in overall population mutation rate and fast production of new alleles (Reich & Goldstein, 1998). Such a signature can then be detected, many generations later, by the population distribution of the individual distance between allele pairs. The *k* test developed by Reich & Goldstein (1998) is constructed to test the null hypothesis that alleles have emerged randomly throughout the population's history, and not concentrated in a particular point in time (a population expansion). Similarly, under the population expansion scenario, all loci are expected to have expanded at the same time and thus, the same authors devised another statistic, *g*, aimed at assessing the congruence of the calculated age of the most ancient allele bifurcation of all loci (Reich & Goldstein, 1998). The estimated *g* can then be compared against a table of simulated cut-off values to assess its significance (Reich *et al.*, 1999). The Microsoft Excel macro KGTESTS (Bilgin, 2007) was employed to estimate single locus *k*, multilocus *g*, and the multilocus associated probability of *k*. Samples were tested independently as structuring can increase the signal of expansion (Reich & Goldstein, 1998). As *DAC1-35* carried strong negative *k* values, but was also a

candidate for null alleles, *g* values and probabilities of *k* were calculated with and without loci affected by null alleles (*DAC1-35* and *DAC5-70*).

Sudden reductions in population size, commonly known as bottlenecks, also leave an imprint on allelic composition. As the effective size of a population diminishes, rare alleles are lost faster than they are replaced by mutation and allelic diversity diminishes (Spencer *et al.*, 2000; England *et al.*, 2003), but as rare alleles have little effect on heterozygosity, the latter is expected to remain unaffected for a number of generations (2 – 4 N_e) resulting in an heterozygosity excess with respect to the number of alleles (Cornuet & Luikart, 1996). The heterozygosity, measured as gene diversity (H_E), can then be compared with that expected in a population of constant size with the observed number of alleles, the heterozygosity at mutation drift equilibrium (H_{eq}), assuming particular mutation models: Infinite Allele Model (IAM), Stepwise Mutation Model (SMM), or a combination of both, the two phase model (TPM). Note that such a test evaluates an excess of heterozygosity ($H_E > H_{eq}$) not excess of heterozygotes ($H_O > H_E$). The probability of H_E being within the 95% CI of H_{eq} was assessed with BOTTLENECK 1.2.02 (Piry *et al.*, 1999), under the three mutation models (IAM, SMM, and TPM: 95% SMM, variance=10) and with a 1000 simulation iterations. As loci deviating from the HWE can potentially affect bottleneck signals (Luikart & Cornuet, 1998), those with evidence of null alleles were not included for the analysis (*DAC1-35*, *DAC5-70*).

7.4 Results:

7.4.1 Quality assurance, locus characteristics and conformity to expectations:

The genotyping error rate was low with 98.4% of alleles identical in both amplifications of the duplicated sample, and most of the errors were at two loci: DAC1-35 (92% accuracy) and DAC5-70 (96% accuracy).

The number of alleles per locus ranged between 8 and 56, and the observed heterozygosity per locus between 0.089 and 0.946. The mean observed heterozygosity across loci within samples was between 0.666 and 0.728 (Table 6.2). The exclusion of *DAC1-35* and *DAC5-70* did not significantly alter the mean observed heterozygosity. Two markers, *DAC1-35* and *DAC5-70*, showed deviations from the HWE in the form of heterozygote deficiency. MICRO-CHECKER suggested the presence of null alleles in some of the samples in some of the years for *DAC1-35* and for all samples in all years for *DAC5-70*. No sample-locus combination

showed evidence of stuttering errors or large allele dropout. Once the two markers outside HWE were removed, all samples complied with HWE expectations.

No evidence of linkage disequilibrium was found among any combination of loci when all samples were pooled together. However, significant linkage ($p<0.001$) was found in up to three genotype comparisons in each year, but none were consistent across samples or years. The source of the disequilibrium was always one or two individuals showing rare alleles or unusual allele combinations at two or more loci for the particular sample they were collected in. The genotypes of these individuals were checked several times, and no evidence of allele miscoring, cross-colour pull-ups, or size-standard miscalling was found, suggesting the genotypes are true. The source of the linkage disequilibrium was probably due to migrants or rare alleles rather than actual genomic linkage among loci.

7.4.2 Testing the existence of differentiation among samples and pairwise compassions:

7.4.2.1 Simulated power of the microsatellite loci to detect differentiation:

POWSIM suggested that the power of the loci suite was 99% to detect differentiations as low as 0.0025 with sample sizes of 100 individuals, or as low as 0.005 for sample sizes of 50 individuals.

7.4.2.2 Testing for the existence of structure among samples:

When all samples were pooled each year, HW became highly significant, even after removal of *DAC1-35* and *DAC5-70*, suggesting that all samples in each year do not belong to the same panmictic population. The patterns of genic subdivision suggested that the allele distribution is consistently significantly different from random ($p<0.001$) for a number of loci (*DAG4-64, DAC3-14, DAC4-40, DAC1-90, DAC5-70*), further suggesting the existence of genetic substructure. When studied in more detail it became clear that there were marked and consistent differences in allele frequencies for some loci between some locations. The global Θ_{WC} (Weir & Cockerham, 1984), which measures the level of population substructuring, including all samples and markers was 0.004 (99% CI 0.001-0.008), suggesting weak structuring of the dab samples. Pairwise estimates of population differentiation (Table 7.1) ranged from negative values (=0, indicating no structure) to a maximum Θ_{WC} of 0.019 between OfF07 and RwB07, and a maximum Θ'_{WC} of 0.083 between NeD05 and StB05. In general, Θ_{WC} and Θ'_{WC} values were close to zero and non significant between samples from the same sea basin (i.e. North Sea, Irish Sea), but higher and highly significant ($p<0.001$) when

comparing samples across sea basins. The pattern between North Sea and Irish Sea was repeated every year. In contrast, the English Channel samples' significant Θ_{WC} values with other basins changed across years. The two Atlantic samples of 2006 displayed highly significant ($p<0.001$) genetic differentiation against all North Sea samples, but were undifferentiated from Irish Sea samples based on Θ_{WC}.

The K-Means clustering analysis suggested that k=2 best fitted the data for all years. Sample membership for k=2, k=3 and k=4 were also calculated. For k=2, all North Sea samples together with RyE and LyB formed one cluster in all years with the exception of OfF05 which clustered with Irish Sea samples. Irish Sea, including Cardigan Bay and Atlantic samples, formed a second cluster in all years. Although the most likely (and temporally stable) k value was two, the patterns of subdivision between samples for larger k's can give insights into further dissimilarities between samples. When three clusters were forced, OfF05 and LyB05 composed a separate group from North Sea and Irish Sea in 2005; in 2006 the imposition of three and four clusters separated Atlantic and Irish Sea (South) samples from the rest of Irish Sea and OfF06 and InF06 separated from the rest of North Sea. Increasing k to three and four in 2007 separated RwB07 and LyB07 from Irish Sea and North Sea respectively; while in 2008, k of three or four resulted in the segregation of InC08 and LyB08 from Irish Sea and North Sea.

Table 7.1: Pairwise estimates of differentiation (Θ_{wr} & Θ'_{wr}) of dab, *Limanda limanda*, around the British Isles

Pairwise estimators of differentiation between samples, sorted by years: 2005-2008; Tables correspond to pairs of Θ_{wc} (above diagonal) and heterozygosity corrected Θ'_{wc} (below diagonal). Significiance values are denoted with stars (*=below 0.05; **=below 0.01; ***=below 0.001). The values have been shaded in orage and yellow for ease of interpretation: The stronger the colour (orange) the higher the relative value among comparisons (within estimator and year). The lines represent the borders between samples in different basins: North Sea (top and left), Irish Sea, and Atlantic (bottom and right)

2005	NeD05	AmB05	OfF05	LyB05	InC05	StB05
NeD05		-0.002	0.004 *	0.004 **	0.007 ***	0.012 ***
AmB05	0.000		0.004 *	0.004	0.004	0.008 **
OfF05	0.042	0.030		0.005	0.006	0.011 ***
LyB05	0.019	0.015	0.028		0.003 *	0.005 *
InC05	0.042	0.024	0.008	0.030		0.003 ***
StB05	0.083	0.045	0.037	0.039	0.044	

2006	NeD06	AmB06	OfF06	InF06	RyE06	LyB06	CeS06	ScB06	InC06	RwB06	LiV06	StB06	DuB06	NoI06	WoI06
NeD06		0.002 ***	0.002	0.002 **	0.002 ***	0.001	0.007 ***	0.007 ***	0.008 ***	0.003	0.004 ***	0.006 ***	0.005 ***	0.004 ***	0.008 ***
AmB06	0.027		0.003	0.001	0.002	0.003	0.008 ***	0.008 ***	0.007 ***	0.004 **	0.007 ***	0.004 ***	0.007 ***	0.004 ***	0.010 ***
OfF06	0.018	0.009		0.002	0.009 ***	0.009 *	0.015 ***	0.018 ***	0.013 ***	0.010	0.010 ***	0.013 ***	0.014 ***	0.013 ***	0.016 ***
InF06	0.022	0.029	0.015		0.003 ***	0.004 **	0.009 ***	0.010 ***	0.008 ***	0.007 ***	0.007 ***	0.008 ***	0.009 ***	0.005 ***	0.009 ***
RyE06	0.014	0.014	0.025	0.028		0.002 ***	0.007 ***	0.007 ***	0.008 ***	0.004 ***	0.006 ***	0.004 ***	0.004 ***	0.004 ***	0.009 ***
LyB06	0.010	0.022	0.036	0.030	0.010		0.002 ***	0.002 *	0.002 *	-0.001	0.001	0.003 *	0.002 **	0.002 **	0.003 ***
CeS06	0.032	0.051	0.056	0.058	0.032	0.023		-0.001	0.000	0.002	0.001	0.003	0.001	0.001 *	0.001 *
ScB06	0.022	0.049	0.063	0.060	0.023	0.012	0.001		-0.001	0.002	0.001	0.001	0.001	0.001 *	0.002 *
InC06	0.048	0.044	0.048	0.057	0.032	0.023	0.017	0.003		0.003	0.003	0.003 *	0.001	0.003 **	0.001
RwB06	0.010	0.016	0.019	0.043	0.009	-0.010	0.001	0.000	-0.001		0.002	0.001	0.000	0.001	0.004
LiV06	0.018	0.040	0.057	0.061	0.022	0.012	0.010	-0.005	0.026	-0.003		0.003	0.000	0.004 *	0.002 *
StB06	0.022	0.020	0.035	0.052	0.008	0.014	0.012	-0.003	0.012	-0.009	-0.004		0.003	0.002	0.004 ***
DuB06	0.024	0.039	0.044	0.067	0.017	0.026	0.011	0.008	0.009	-0.003	0.008	0.011		0.001	0.002 *
NoI06	0.022	0.025	0.054	0.031	0.022	0.019	0.007	0.010	0.022	0.002	0.022	0.009	0.007		0.002
WoI06	0.049	0.045	0.062	0.046	0.031	0.024	0.015	0.022	0.026	0.016	0.031	0.012	0.030	0.008	

Table 7.1 (Cont.)

2007	NeD07	AmB07	OfF07	InF07	RyE07	LyB07	InC07	RwB07	LiV07	StB07
NeD07		0.002	0.000	0.001	0.002 ***	0.001	0.004 ***	0.011 ***	0.005 ***	0.005 ***
AmB07	0.000		-0.001	-0.003	0.000	0.002	0.006 ***	0.016 ***	0.008 ***	0.009 ***
OfF07	0.001	0.005		-0.003	0.001	0.001	0.007 ***	0.020 ***	0.009 ***	0.011 ***
InF07	-0.001	-0.007	-0.012		0.000	0.003	0.006 ***	0.016 ***	0.007 ***	0.009 ***
RyE07	0.012	0.000	0.005	-0.002		0.000	0.003 *	0.011 ***	0.003 ***	0.004 ***
LyB07	0.000	-0.004	-0.006	0.013	-0.004		0.005	0.007 ***	0.001	0.003
InC07	0.027	0.030	0.036	0.041	0.012	0.002		0.007 **	0.002	0.000
RwB07	0.057	0.060	0.068	0.055	0.054	0.028	0.026		0.002 **	-0.001
LiV07	0.032	0.037	0.047	0.033	0.026	0.001	0.011	0.009		-0.001
StB07	0.038	0.047	0.055	0.049	0.026	0.009	-0.003	0.009	-0.006	

2008	NeD08	AmB08	OfF08	InF08	RyE08	LyB08	InC08	LiV08	StB08
NeD08		0.002	0.001 *	0.005 **	0.004 *	0.003	0.021 ***	0.011 ***	0.011 ***
AmB08	0.001		0.001	0.001	0.002 **	0.001	0.015 ***	0.006 ***	0.009 ***
OfF08	0.009	0.001		0.001	0.000	0.004 *	0.017 ***	0.010 ***	0.007 ***
InF08	0.013	0.002	0.004		0.000	0.004	0.017 ***	0.007 ***	0.007 ***
RyE08	0.010	0.005	0.003	-0.004		0.003	0.013 ***	0.009 ***	0.006 ***
LyB08	0.011	0.013	0.025	0.006	0.006		0.007	0.004 *	0.006 ***
InC08	0.047	0.053	0.042	0.049	0.030	0.023		0.001	0.003
LiV08	0.044	0.044	0.047	0.038	0.043	0.035	0.002		0.002
StB08	0.058	0.061	0.055	0.049	0.047	0.049	0.016	-0.002	

The AMOVA (Table 7.2) suggested that around 98.4% of the variation was within individuals, while 0.8, 0.1 and 0.7% were among individuals within populations, among populations within clusters, and between clusters respectively. Despite the low levels, all were significant (p=0.036, p=0.009, and p<0.001). When φ_{ST} was corrected for high heterozygosity, differentiation among populations within clusters remained low (φ'_{ST} =0.004), but became more evident between clusters (up to φ'_{ST}=0.024 in 2008).

In the neighbour-joining trees North Sea samples clustered together at one side of the tree, while Irish Sea samples clustered at the other end in all years (Figure 7.2). One exception was the sample OfF05 (N=21), which clustered with the Irish Sea samples. RyE Bay was generally incorporated within the North Sea cluster. On the other hand, the samples from Lyme Bay always fell in between the North Sea and Irish Sea samples. In 2006, the Atlantic samples clustered together with Irish Sea samples.

7.4.3 Geo-referenced Analysis:

7.4.3.1 Relationship between genetic distance and geographic distance:

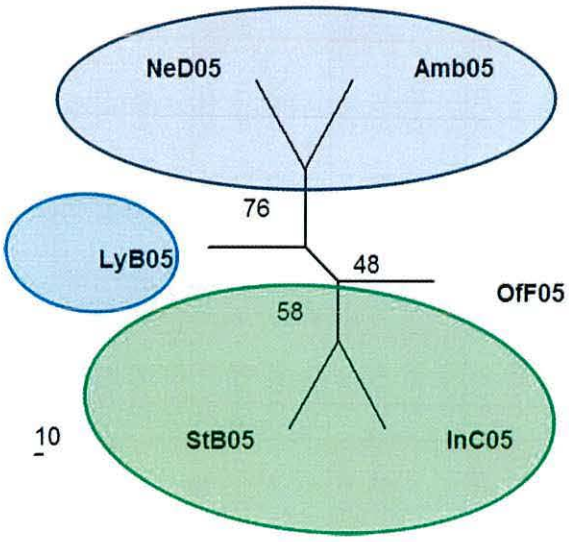
Pairwise $\Theta_{WC}/(1-\Theta_{WC})$ among samples were significantly correlated with the shortest sea distance (km) between samples for 2006 (Mantel test, r=0.700, p<0.001), 2007 (r=0.641, p<0.001) and 2008 (r=0.816, p<0.001). The relationship was weaker for samples in 2005 (r=0.424, p=0.057) (Table 7.3). Using the heterozygosity standardised version, $\Theta'_{WC}/(1-\Theta'_{WC})$, did not alter markedly the significance values and increased only slightly the slope of the relationships (Figure 7.3). However, when sea basin was accounted for (partial mantel test), all values became non-significant (except for 2006 when including Atlantic samples), suggesting that the function "sea basin" performs better at explaining genetic distance than geographic distance *per se* among samples. All test of isolation by distance within basin were not significant, reinforcing the role of basin as a primary factor impacting the distribution of genetic diversity of dab. Finally, if geographic distance was divided into three classes: below 400 km (representing within basin), 400 to 1000 km (distance from either North Sea or Irish Sea to English Channel), and above 1000 km (between Irish Sea and North Sea), the slopes of the relationship between genetic distance and geographic distance became nearly horizontal and non-significant, confirming the initial pattern of isolation by distance truly represents the genetic distance associated with the different basins, with LyB samples falling in between both basins.

Table 7.2: AMOVA of dab, *Limanda limanda*, around the British Isles

Analysis of Molecular Variance of dab samples around the British Isles for four years (**2005-2008**) and overall (**All**). **Sources of variation**: Individuals; Individuals within Samples; Samples within Clusters; and Clusters. Cluster are those estimated by the K-means clustering analysis and correspond to North Sea-English Channel and Irish Sea basins. **%var**= percentage of variation explained by the source of variation. φ_{ST} = estimated amount of differentiation in the system in each source of variation. **St.err**= Standard error around the estimated φ_{ST}. **p-value**= probability that the 95% CI of the estimated φ_{ST} encompasses zero. φ'_{ST}: estimated amount of differentiation in the system after correction for heterozygosity. The analysis was repeated with only **Females** and only **Males**. The amount of genetic variation explained by the different **years** was also tested with AMOVA, using year as the cluster (**All by years**).

			Whole					Females					Males				
	Source of Variation	**Nested in**	**%var**	φ_{ST}	**st.err.**	**p-value**	φ'_{ST}	**%var**	φ_{ST}	**st.err.**	**p-value**	φ'_{ST}	**%var**	φ_{ST}	**st.err.**	**p-value**	φ'_{ST}
2005	Within Individual	-	0.981	0.019	0.011	--	--	0.979	0.021	0.019	--	--	0.980	0.020	0.007	--	--
	Among Individual	Samples	0.012	0.012	0.008	0.040	--	0.011	0.011	0.015	0.151	--	0.014	0.014	0.007	0.090	--
	Among Samples	Clusters	0.003	0.003	0.002	0.006	0.010	0.006	0.006	0.003	0.012	0.020	0.004	0.004	0.003	0.036	0.014
	Among Clusters	-	0.004	0.004	0.002	<0.001	0.013	0.004	0.004	0.005	<0.001	0.015	0.003	0.003	0.003	0.102	0.009
2006	Within Individual	-	0.986	0.014	0.008	--	--	0.991	0.009	0.010	--	--	0.983	0.017	0.010	--	--
	Among Individual	Samples	0.008	0.008	0.006	0.013	--	0.004	0.004	0.007	0.224	--	0.008	0.009	0.009	0.054	--
	Among Samples	Clusters	0.001	0.001	0.000	<0.001	0.005	0.002	0.002	0.001	0.015	0.006	0.002	0.002	0.001	0.017	0.006
	Among Clusters	-	0.004	0.004	0.003	<0.001	0.014	0.003	0.003	0.002	<0.001	0.012	0.007	0.007	0.004	0.001	0.022
2007	Within Individual	-	0.985	0.015	0.010	--	--	0.980	0.020	0.011	--	--	0.997	0.003	0.013	--	--
	Among Individual	Samples	0.009	0.009	0.008	0.008	--	0.015	0.015	0.008	0.001	--	-0.004	-0.004	0.011	0.705	--
	Among Samples	Clusters	0.000	0.001	0.001	0.088	0.002	0.001	0.001	0.001	0.081	0.003	0.000	0.000	0.002	0.414	0.001
	Among Clusters	-	0.005	0.005	0.003	0.005	0.018	0.005	0.005	0.003	<0.001	0.016	0.007	0.007	0.004	<0.001	0.022
2008	Within Individual	-	0.984	0.016	0.008	--	--	0.981	0.019	0.009	--	--	0.987	0.013	0.010	--	--
	Among Individual	Samples	0.008	0.008	0.007	0.036	--	0.009	0.009	0.007	0.072	--	0.005	0.005	0.010	0.241	--
	Among Samples	Clusters	0.001	0.001	0.001	0.009	0.004	0.001	0.001	0.001	0.091	0.003	0.002	0.003	0.001	0.020	0.009
	Among Clusters	-	0.007	0.007	0.003	<0.001	0.024	0.009	0.009	0.004	0.172	0.030	0.005	0.005	0.002	0.013	0.018
All	Within Individual	-	0.985	0.015	0.009	--	--										
	Among Individual	Samples	0.009	0.009	0.006	<0.001	--										
	Among Samples	Clusters	0.001	0.001	0.000	<0.001	0.004										
	Among Clusters	-	0.005	0.005	0.003	<0.001	0.017										
All by Years	Within Individual	-	0.987	0.013	0.007	--	--										
	Among Individual	Samples	0.009	0.009	0.006	<0.001	--										
	Among Samples	Years	0.004	0.004	0.002	<0.001	0.013										
	Among Years	-	0.000	0.000	0.000	<0.001	0.000										

2005



2006

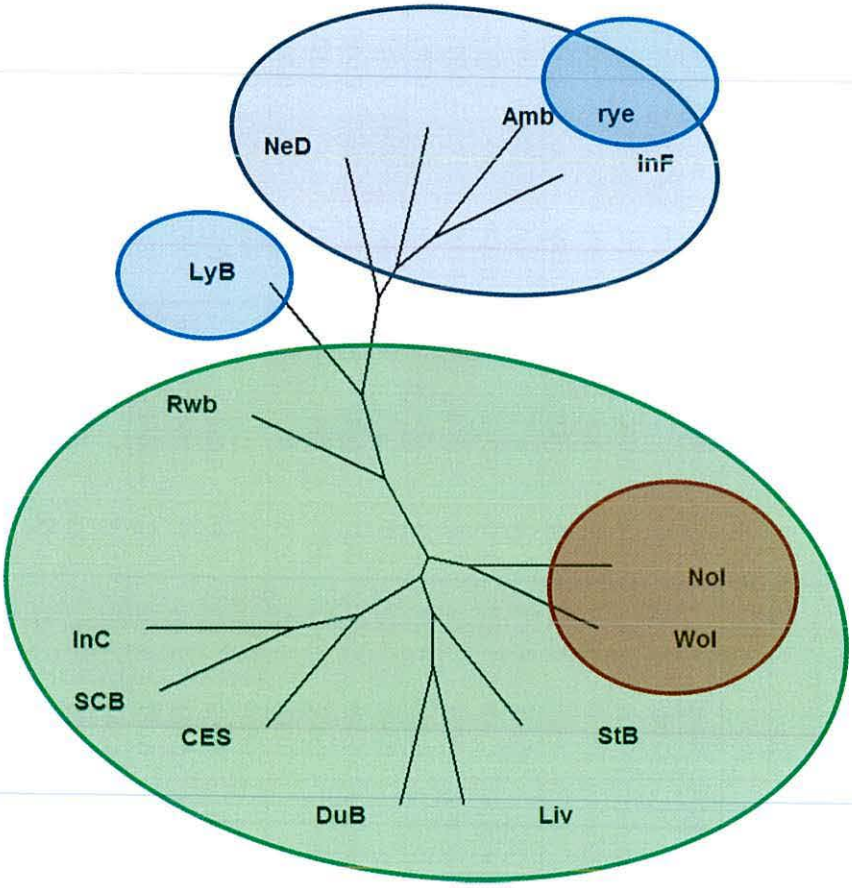
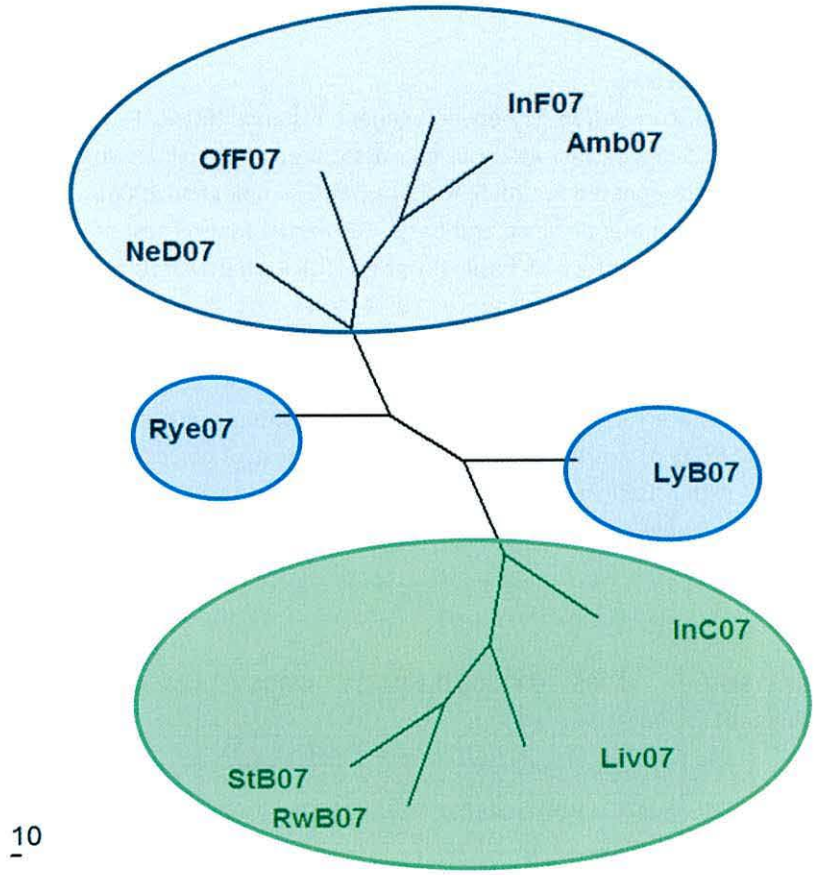


Figure 7.2: Neighbour joining trees.
To understand the relationships between samples, neighbour-joining trees based on Nei's D_S (Nei, 1972). There is one for each year (2005-2008). Robustness of the nodes was assessed with a 100 bootstraps, the number of times two samples clustered together out of the 100 bootstraps (bootstrap support value) is given near the node. Coloured bubbles have been placed to ease understanding: Dark blue represents samples collected from the North Sea; Light blue are samples collected in the English Channel; Light Green are those samples collected in the Irish Sea; and brown is reserved for those samples collected in the Atlantic coast of Ireland.

2007



2008

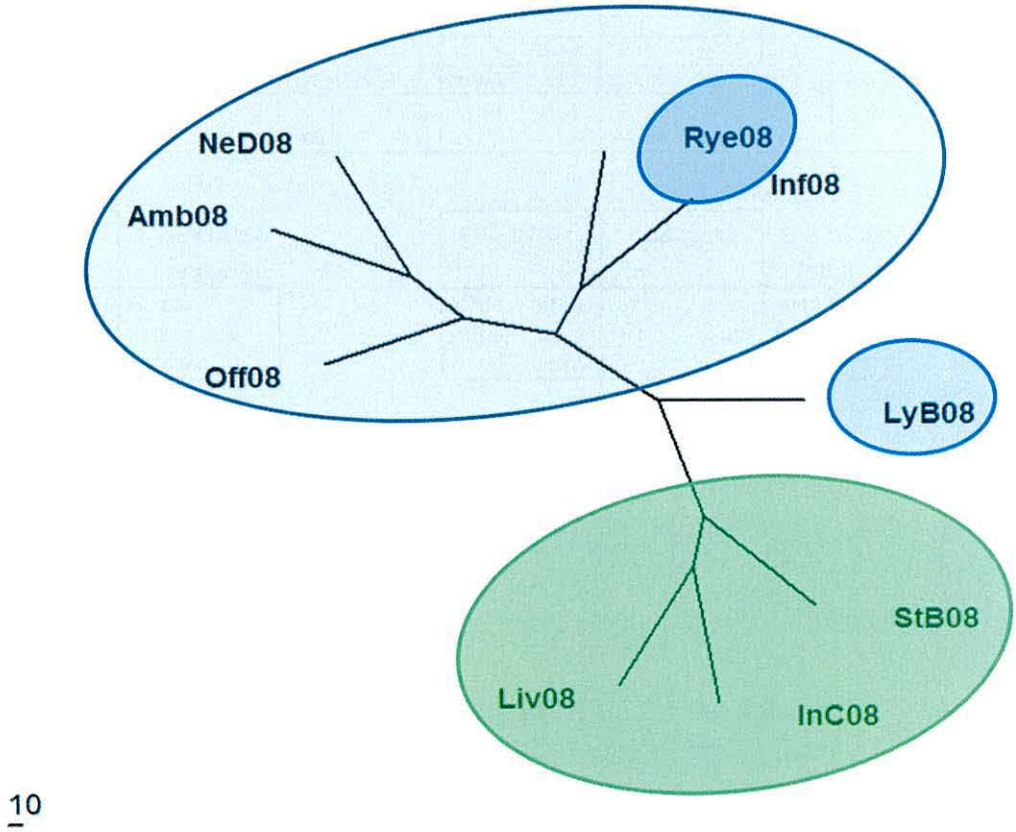


Figure 7.2: (Cont.)

Table 7.3: Mantel tests

Mantel test of correlation (r) between genetic distance (Fst/(1-Fst) and geographic distance (log) and basin (North Sea-Irish Sea-Atlantic) and associated probabilities (p). Tests performed for each year separately, and repeated for 2005 (without Off05 sample) and 2006 (without the Atlantic samples). (**A**) Correlation of genetic distance and basin. (**B**) Partial Mantel test of genetic and geographic distance, controlling for the effect of basin (In grey). (**C**) Partial Mantel test of genetic distance and basin, controlling for geographic distance. (**D**) Mantel test of genetic and geographic distances. Two estimators of genetic distance were employed: ΘWC (**ANALYSIS 1**) and its heterozygosity corrected form, Θ'WC (**ANALYSIS 2**). Significant probability values after Bonferroni correction in bold. Bottom left (**ANALYSIS 3**) are the results of the Mantel test within sea basin for each year (2006-2008). Bottom right (**ANALYSIS 4**) are the results of the Mantel test of correlation between genetic and geographic distances within basin (<400km), between either North Sea or Irish Sea and English Channel (400-1000km), and between North Sea and Irish Sea (>1000km).

ANALYSIS 1: log(geo)		2005		2005 (no Off)		2006		2006 (no AT)		2007		2008	
		r	p	r	p	r	p	r	p	r	p	r	p
$\frac{\Theta_{WC}}{1-\Theta_{WC}}$	A	0.521	0.067	0.627	**0.000**	0.603	**0.000**	0.738	**0.000**	0.729	0.005	0.908	**0.000**
	B	-0.065	0.557	0.103	0.341	0.497	**0.000**	0.304	0.011	0.170	0.133	0.373	0.049
	C	0.424	0.065	0.478	0.074	0.238	0.035	0.388	0.005	0.478	0.005	0.741	0.011
	D	0.340	0.073	0.471	0.067	0.701	**0.000**	0.717	**0.000**	0.641	**0.000**	0.816	**0.000**

ANALYSIS 2: log(geo)		2005		2005 (no Off)		2006		2006 (no AT)		2007		2008	
		r	p	r	p	r	p	r	p	r	p	r	p
$\frac{\Theta'_{WC}}{1-\Theta'_{WC}}$	A	0.347	0.134	0.546	0.100	0.602	**0.000**	0.653	**0.000**	0.762	**0.000**	0.921	0.012
	B	-0.183	0.768	-0.016	0.497	0.482	**0.000**	0.428	0.002	0.173	0.130	0.438	0.022
	C	0.366	0.073	0.450	0.093	0.247	0.054	0.177	0.057	0.522	0.005	0.770	0.012
	D	0.133	0.274	0.348	0.125	0.692	**0.000**	0.718	**0.000**	0.664	**0.000**	0.835	**0.000**

ANALYSIS 3: log(geo)		Irish Sea		North Sea	
		r	p	r	p
$\frac{\Theta'_{WC}}{1-\Theta'_{WC}}$	2006	0.396	0.063	0.364	0.350
	2007	0.148	0.416	0.487	0.107
	2008	0.773	0.166	0.014	0.365

ANALYSIS 4: log(geo)	Θ_{WC}		Θ'_{WC}	
	r	p	r	p
<400 km	0.063	0.632	0.170	0.193
400 - 1000 km	0.185	0.151	0.037	0.777
>1000 km	0.160	0.173	0.096	0.417

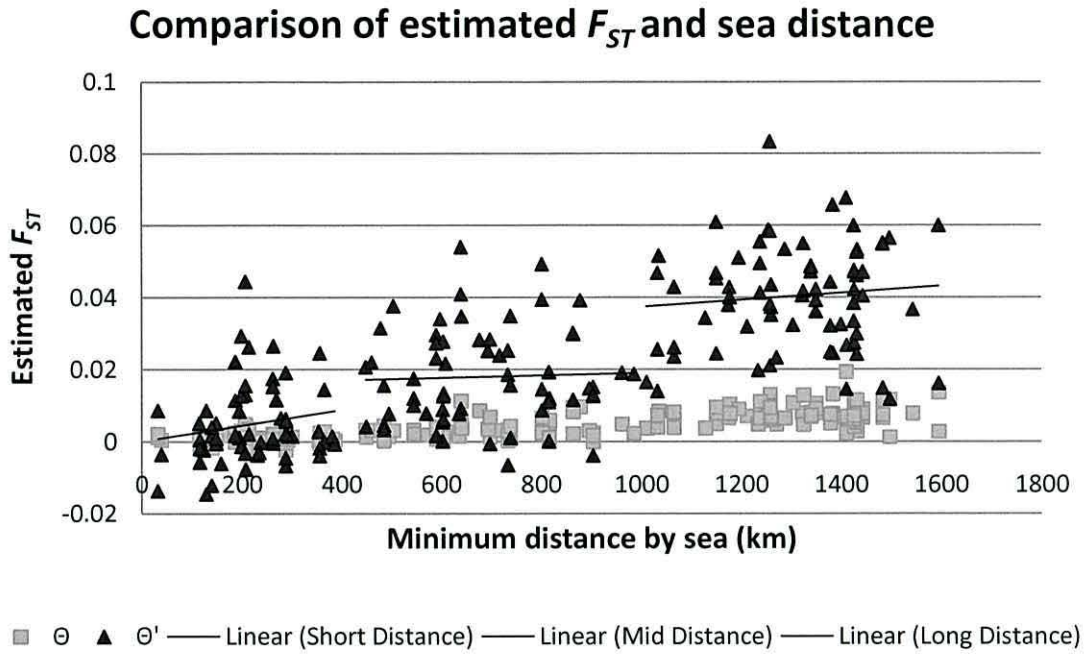


Figure 7.3: Genetic distance (Θ_{WC} and Θ'_{WC}) Vs. Geographic distance (km).

Two estimates of genetic distance ΘWC (light squares) and Θ'WC (corrected for heterozygosity; dark triangles) are compared to minimum distance travelled by sea between two samples. The overall trend is for a positive relationship between genetic and geographic distances; however the relationship is lost (horizontal regression lines) when comparisons are divided into three groups: within basin (short distance), between North and Irish Seas (long distance), or between either seas and the English Channel (medium distance).

7.4.3.2 Evaluation of the most important barriers to gene exchange between locations:

In 2005, uncorrected estimators suggested the barrier between LyB05 and North Sea samples as the most important; however, all corrected estimators positioned the strongest barrier between StB05 and InC05. Bootstrapping of D_S yielded 77.6% support for the StB05-InC05 barrier, with a further 19.6% of barriers lying on the InC05-LyB05, giving weight to the barriers provided by the corrected estimators. When single locus matrices were employed and combined in a single map, it was revealed that the patterns of differentiation were different across loci. The InC05-LyB05 boundary was supported by most loci (5), while the InC05-StB05 and LyB05-NeD05 boundaries were supported by four and three loci respectively.

The larger number of samples in 2006 increased the complexity in the order of importance of the barriers. Θ_{WC} indicated a first boundary between Irish Sea (South) and Irish Sea (North) samples, a second boundary (InF06-RyE06) separated North Sea from English Channel samples, and a third isolated LiV06 from the rest of Irish Sea (North). Θ'_{WC}, G'_{ST_est}, and D_{est}, generated similar first four boundaries, albeit in different orders. To summarise, they

separated North Sea from English Channel across the InF06-RyE06 boundary, and highlighted the LyB06-Irish Sea boundary, effectively segregating the North Sea, English Channel and Irish Sea; they also isolated NeD06 and InC06 from the rest of North Sea and Irish Sea samples. Bootstrap D_S replicates, gave the highest support for the barrier between InC06 and the rest of Irish Sea samples (37%), particularly those to the north (74.5%). Locus wise, the most supported edges, were across the south of Ireland (WoI06-CeS06) and between Irish Sea (South) and Irish Sea (North) (3 loci each), however when 2nd and 3rd order barriers per locus were included, a previously unseen boundary bisected the Irish Sea into East and West, and separated English Channel samples from the Irish Sea. 2006 was the year in which differences in locus barriers for corrected and traditional estimators were most noticeable. The most supported edges when corrected estimates were used highlighted the English Channel-ISS edge, a second boundary snaked through the Irish Sea leaving DuB06 and ISS samples at one side, and StB06 and RwB06 at the other. LiV06 was left isolated.

In 2007, Θ_{WC} and G_{ST_est}, located the first and second barriers north and south of InC07, while the third isolated NeD07. Θ'_{WC}, G'_{ST_est} and D_{est}, also suggested the strongest barrier was north of InC07, but placed the next barrier around RwB07, and a third separated NeD07 and AmB07 from the rest of the North Sea. The InC07-RwB07 boundary was also backed by the highest bootstrap support (76.6%). Interestingly, no locus revealed the InC07-RwB07 edge as most important, which was the case for the InC07-LyB07 edge and the boundaries around NeD07 and OfF07. However, when the second most important barrier in each locus was included, the InC07-RwB07 became the most dominant.

In 2008, BARRIER identified the boundary between LyB08 and InC08 as the strongest for all measures of differentiation (Θ_{WC}, G_{ST_est} , Θ'_{WC}, G'_{ST_est}, and D_{est}). All heterozygosity corrected estimators diagnosed the barrier between InC08 and the other Irish Sea samples as second in importance, while Θ_{WC} and G_{ST_est} suggested the edges around NeD08 as second in importance. The pattern was inverted for the third and fourth barriers. Bootstrap replicates of D_S distances confirmed the edge between InC08 and LyB08 as the most supported (80.6%), followed by the boundary between InC08 and the rest of Irish Sea (17.8%). When only the first barrier for each locus was taken into account, the edge between LiV08 and InC08 was supported by the most loci (5), while StB08-LiV08, StB08-InC08, InC08-LyB08 and edges around NeD08 were supported by three or four loci. However, when the first two barriers per locus were included the edge between InC08 and LyB08 became the most supported (8 loci), suggesting that although locus-specific maximum differences between areas might be more

distributed, the boundary between InC08 and LyB08 reclaims strong cross-boundary differences for much of the genome.

In general there was better agreement in the order of barriers between corrected estimators of differentiation and bootstrapped D_S distances (which are generated by randomly resampling loci with replacement within individuals from the raw data), indicating that corrected estimators are better descriptors of genetic structure and thus of barriers to gene exchanges. Geographically, the eastern part of the English Channel and the south of the Irish Sea seem to be where the strongest and most supported barriers (both by loci and bootstraps) for all years are detected. However, the position of the barriers within the area varied from north of InC in 2005, 2006 and 2007, to south of InC in 2008.

7.4.3.3 Population membership of geo-referenced samples:

GENELAND estimated the most likely number of interbreeding groups of individuals as three (2005) or four (2006, 2007, 2008). The posterior probability maps (Figure 7.4;Figure 7.5;Figure 7.6;Figure 7.7) showed clear subdivision between North Sea and Irish Sea samples for all years. RyE samples normally clustered with North Sea (except 2007), while LyB flipped between North Sea in 2006 and a third group with either RyE (2007) or InC (2005 and 2008). The Atlantic samples of 2006 formed a putative independent breeding unit from the rest.

Figure 7.8The correspondence analysis (CA; Figure 7.8) clearly separated all North Sea samples from Irish Sea-Atlantic samples along the first axis, placing the LyB samples of 2005, 2006, and 2007 in between North Sea and Irish Sea samples. The LyB08 samples clustered with the rest of the North Sea samples. Again, the OfF05 sample clustered with the rest of Irish Sea samples. The sPCA (Figure 7.9) indicated the existence of structure on the first global component (i.e. indicating differences between distant samples) but not the local end of the spectrum (i.e. between neighbouring samples) for all years as seen on the screeplots (sorted eigenvalues) and decomposition values, resulting in two differentiated clusters, one centred in the North Sea (OfF) and another in the southern Irish Sea (South). Samples from LyB were consistently undifferentiated from either North Sea or Irish Sea, suggesting either a mixture of individuals from both locations or the existence of an admixed population in the western part of the English Channel.

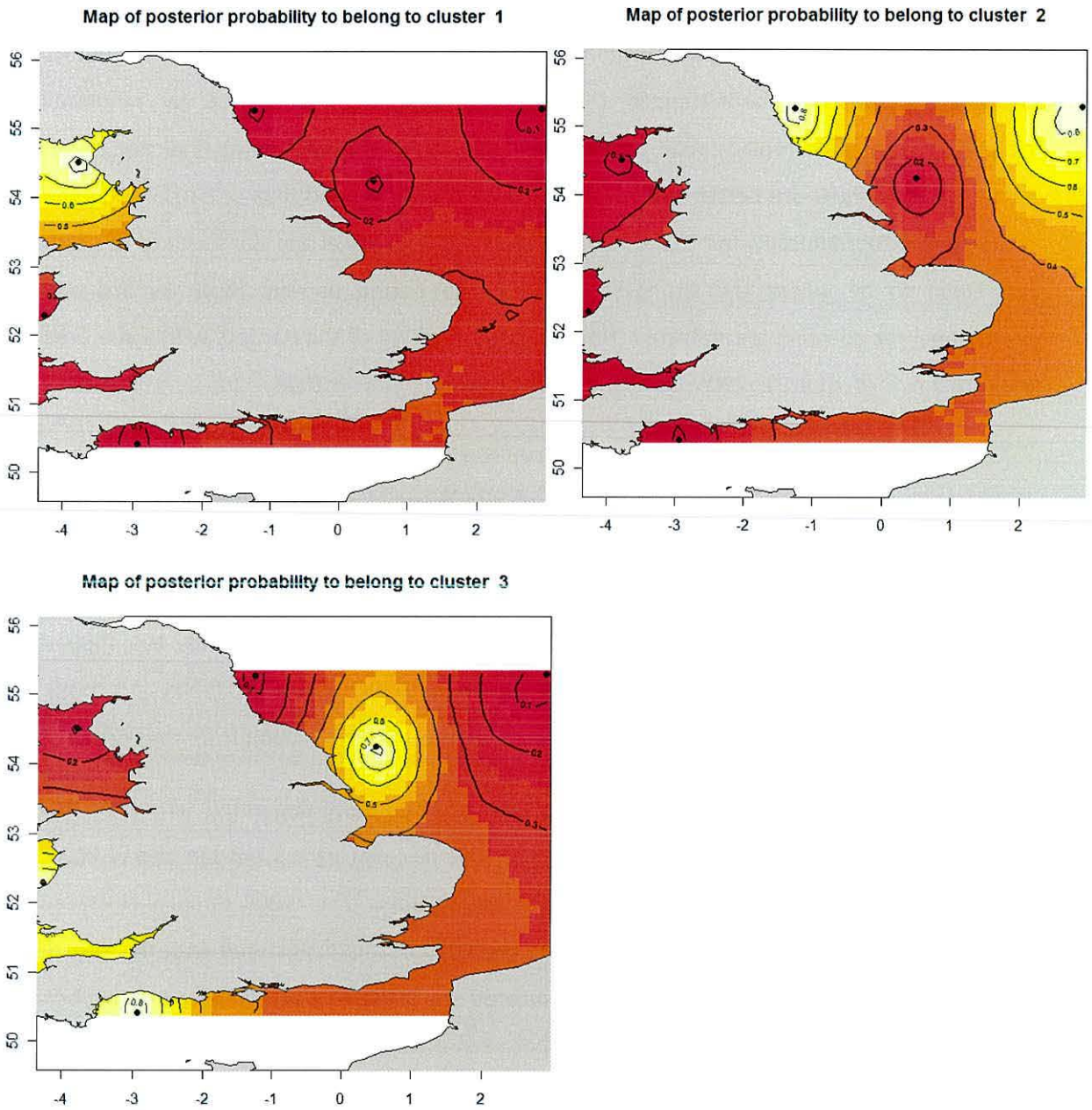


Figure 7.4: GENELAND plots for 2005.

Number of clusters identified by GENELAND from the genotypic data of dab in 2005; Grey represent landmasses (Great Britain, France and Ireland); the samples are identified by small black dots; one plot is created for each identified cluster (clusters 1-4), and the probability of each sample to belong to that cluster is coloured from low (red) to high (yellow and white).

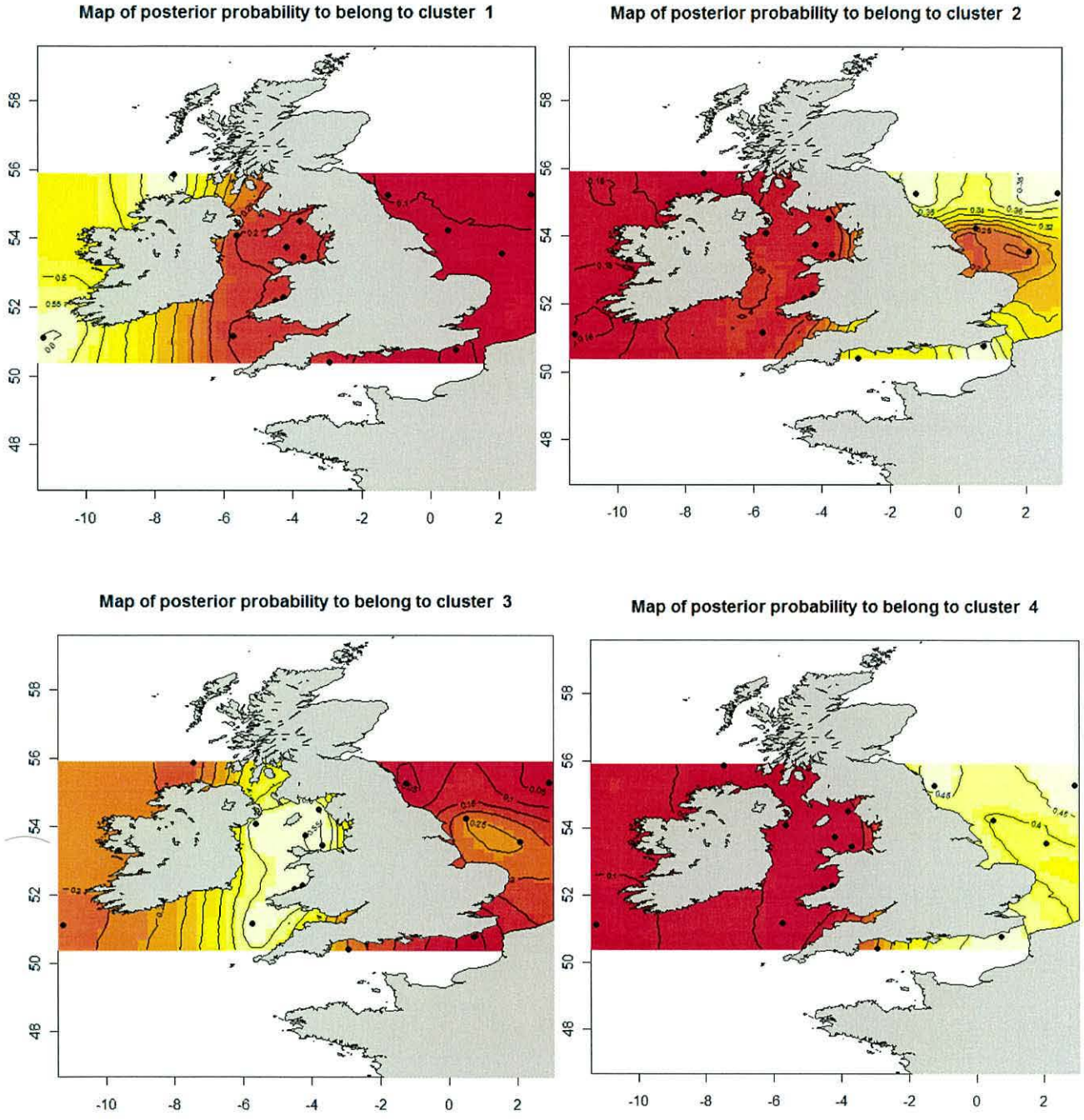


Figure 7.5: GENELAND plots for 2006.

Legend as for the 2005 GENELAND plots.

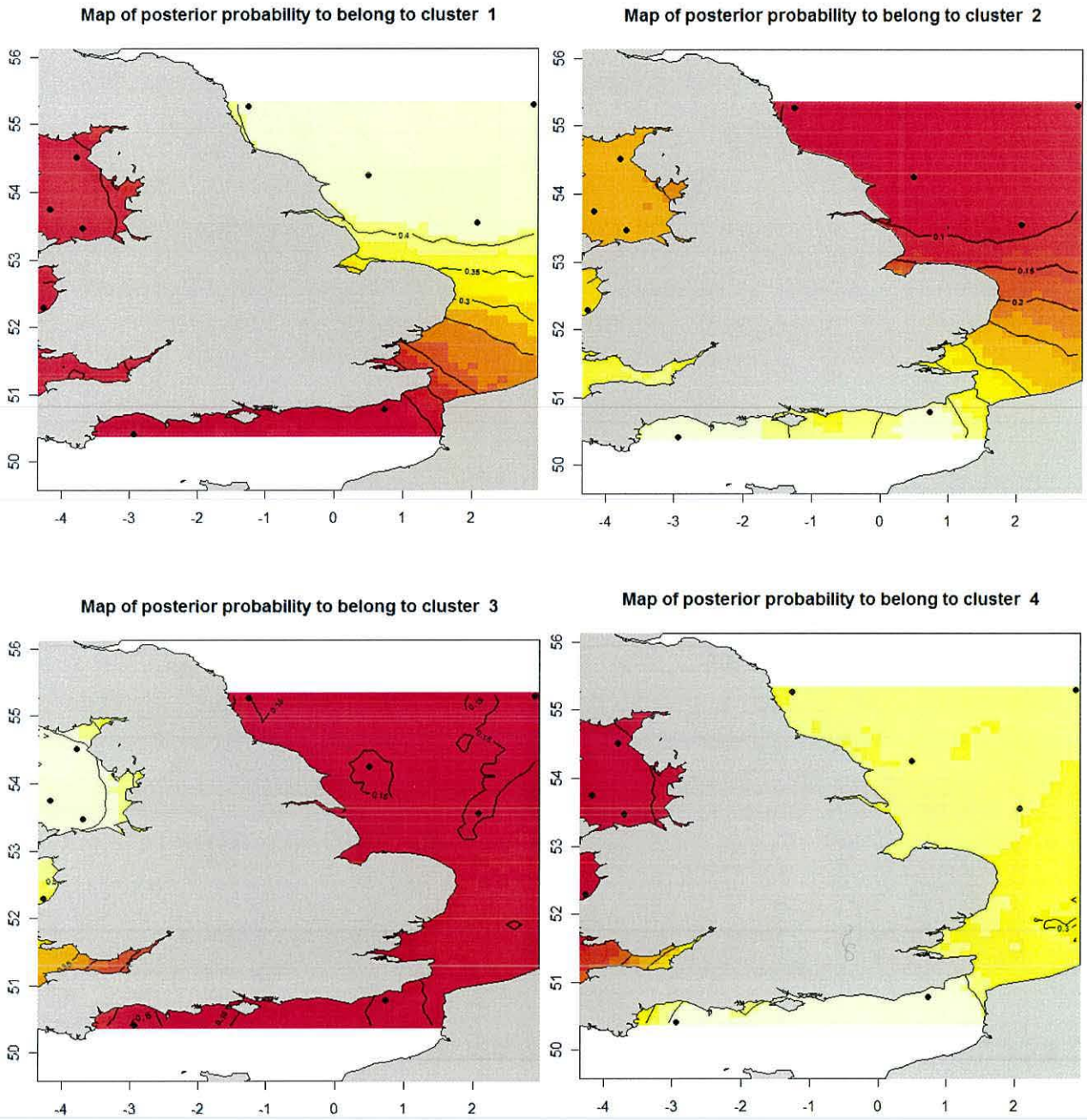


Figure 7.6: GENELAND plots for 2007.

Legend as for the 2005 GENELAND plots.

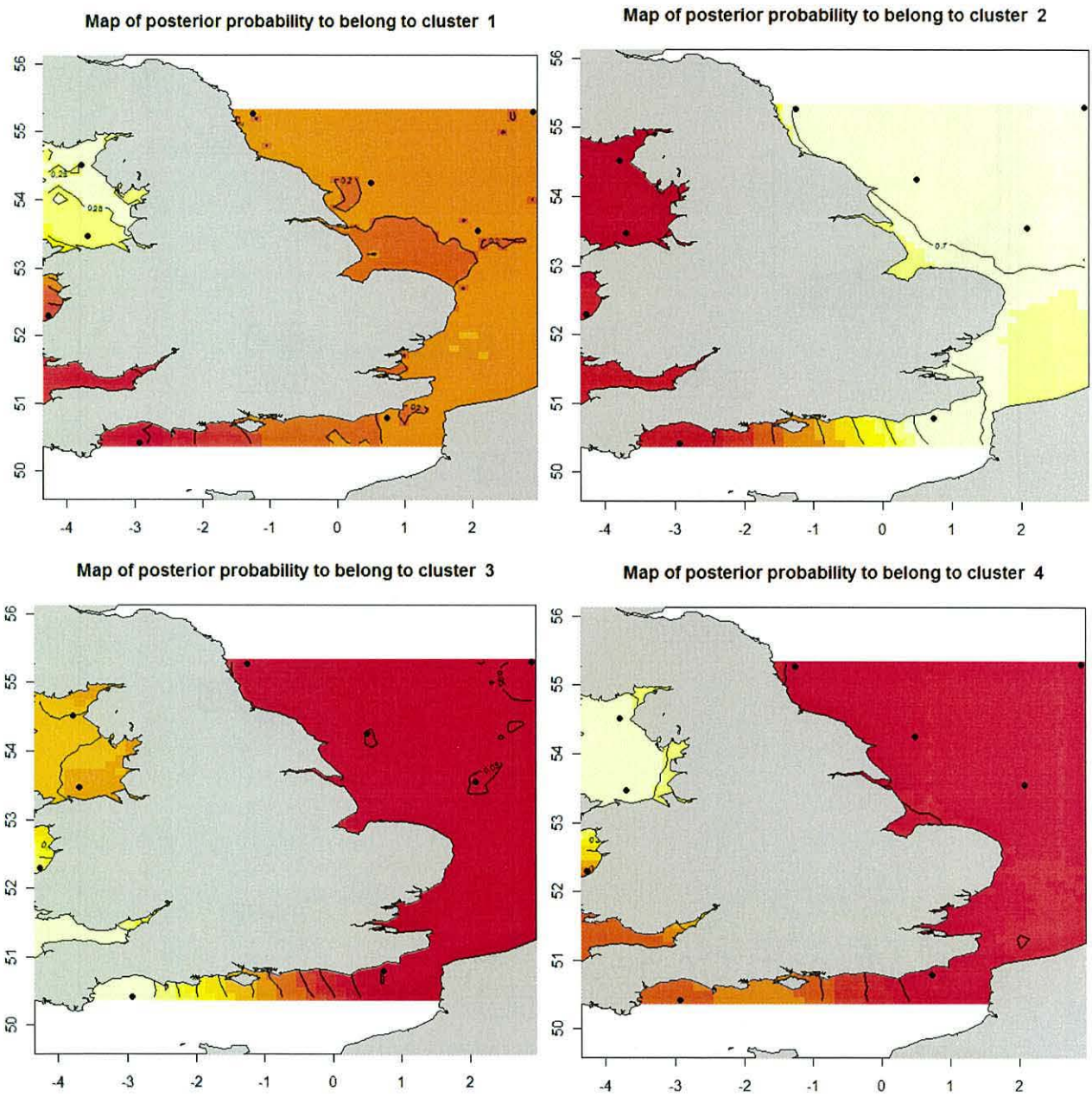


Figure 7.7: GENELAND plots for 2008.

Legend as for the 2005 GENELAND plots.

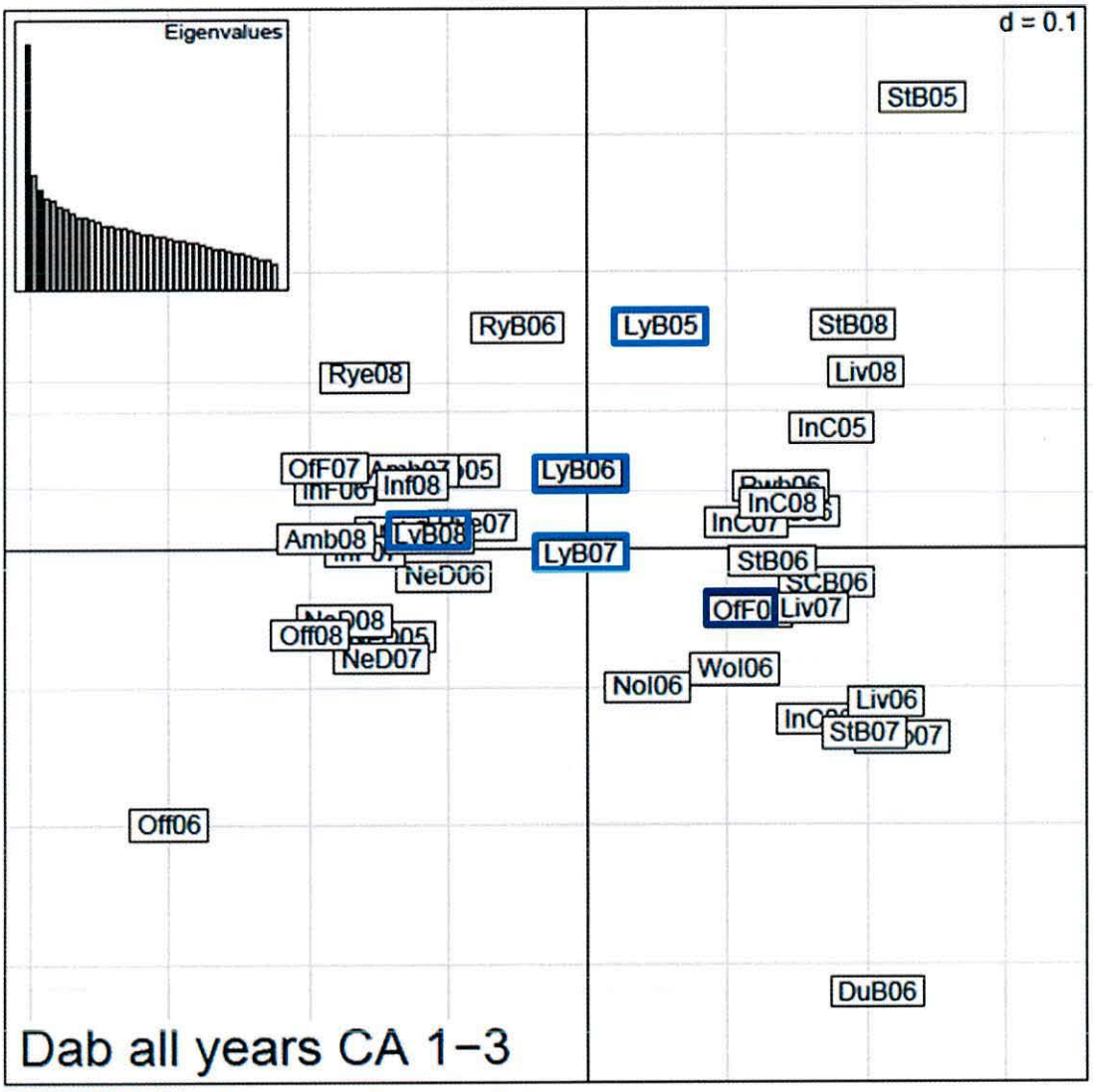


Figure 7.8: Sample correspondence analysis:

Sample correspondence analysis (CA) separating North sea (Left) from Irish Sea (Right) along the first axis (abscissa). Highlighted in light blue are LyB samples. Note that from 2005-2006 LyB are in between Irish and North Seas samples, while in 2008 they clearly cluster with the rest of North Sea. Highlighted in dark blue is the mislabelled OfF05 sample clustering with the rest of the Irish Sea.

2006

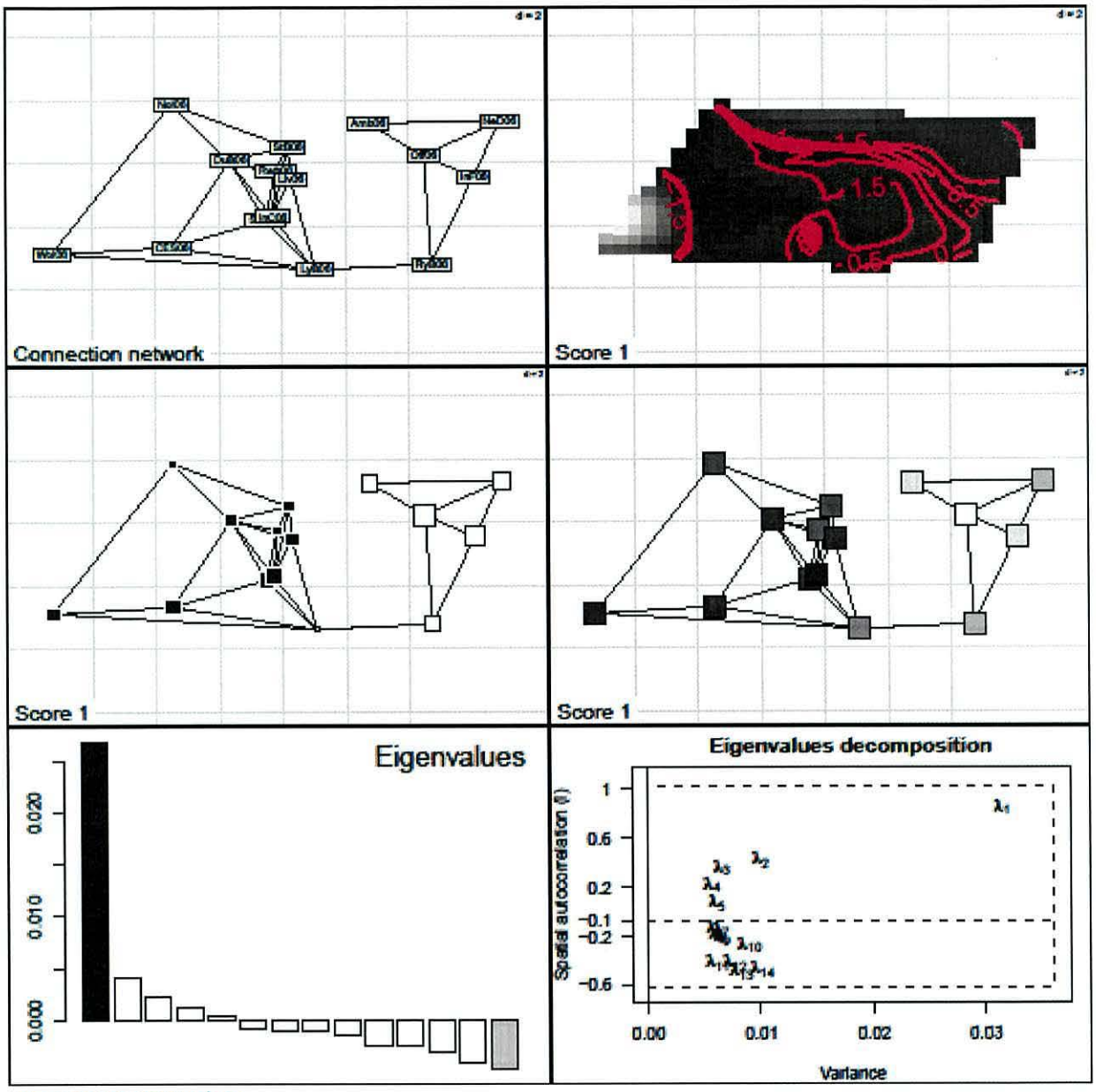


Figure 7.9: sPCA plots.

Spatial Principal Component Analysis of dab samples around the British Isles for the years 2006-2008. The top left plot depicts the connection network used to define spatial weightings. The top right is a representation of sample scores using a local interpolation of the scores with grey levels and contour lines. The two middle plots are also representations of the scores using square size (left) and greyscale (right) to depict similarly or differentiation between samples. The larger (left) or more white/black (right) the square, the more differentiation from squares in the opposite colour. Smaller (left) or greyer (right) squares are less differentiated from the extreme samples. The bottom graphs are displays of eigenvalues: bottom left as screeplots, where values to the left indicate the magnitude of global structures, while values to the right designate local patterns; bottom right as λ values decomposed into variance and spatial autocorrelation components.

2007

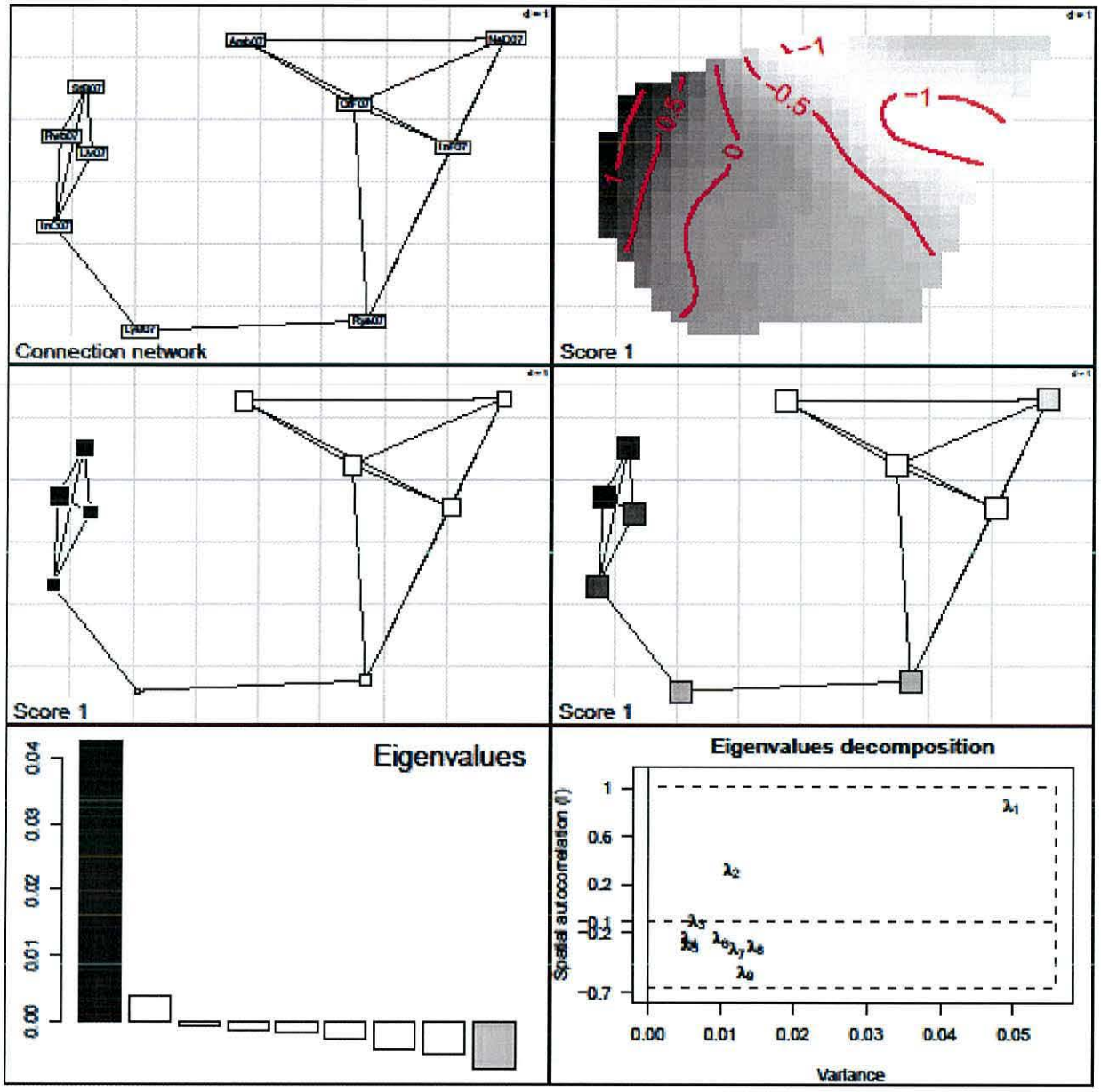


Figure 7.9: (cont.)

2008

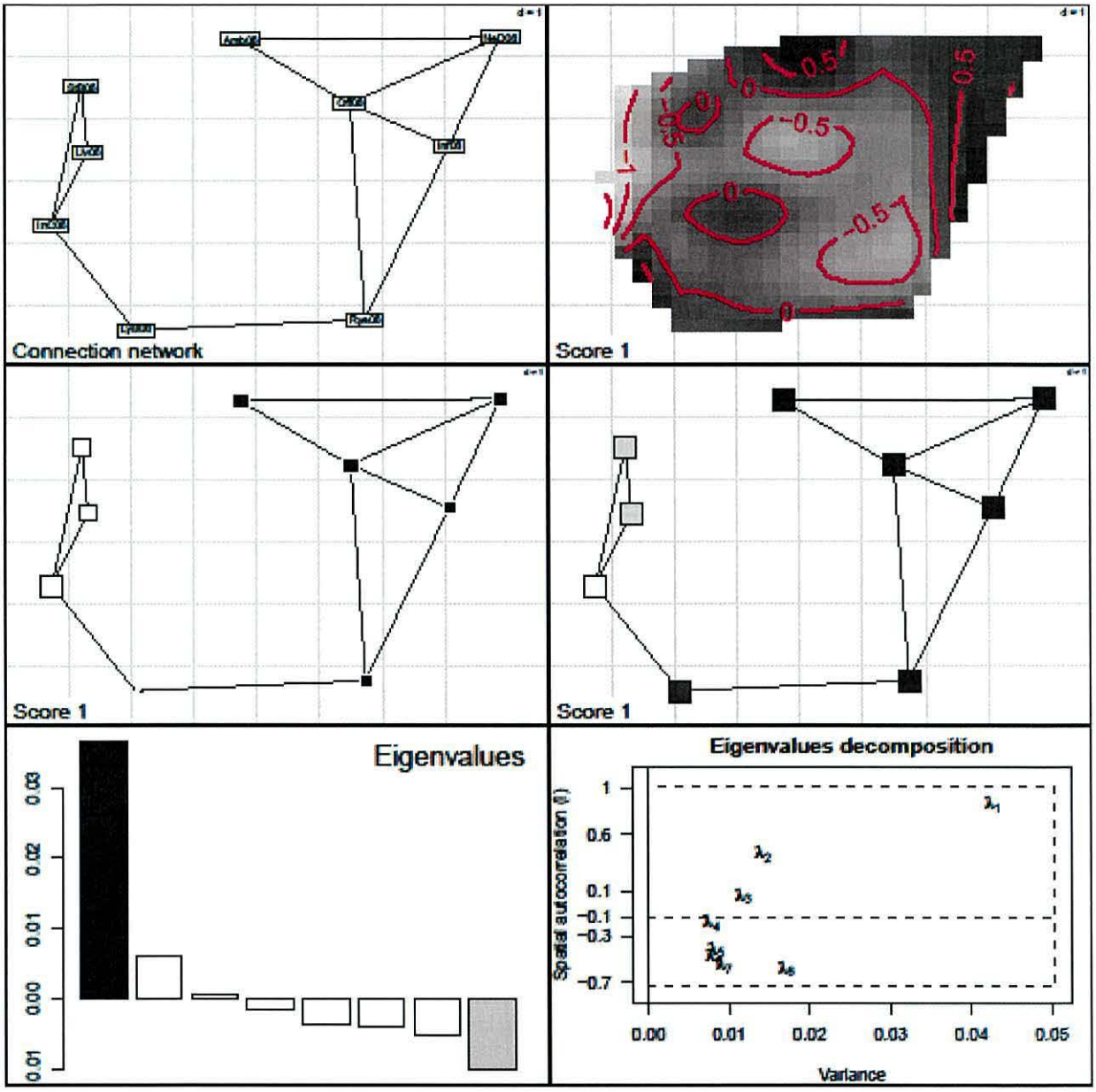


Figure 7.9: (Cont.)

7.4.3.4 Migration and admixture:

Outputs from the individual migrant-detection exercise run in GENECLASS2, whether calculated with the allele frequencies method (Paetkau *et al.*, 1995) or the Bayesian approach (Rannala & Mountain, 1997), were very similar (Table 7.4). However, those individuals identified as migrants by one method not always matched those identified by the other. It was found that the number of discordant individuals (those not identified by both methods) approximately matched those expected by chance (i.e. number of individuals considered x 0.01). Therefore, only those individuals significant with both methods will be discussed.

Several individuals each year had very low probabilities ($p<0.01$) of belonging to the same group of interbreeding individuals as the rest of dab collected in the same location that year. These individuals will be referred to as *migrants* (although other processes could produce such signals, see Discussion). In those cases, GENECLASS2 evaluates the probability of the migrant belonging to any of the other samples included in the analysis, and assigns the individual to whichever is most probable. There were cases in which individuals were assigned to a neighbouring sampling site; and others in which individuals were assigned to another basin; the latter are hereafter referred to as "trans-basin migrants". In 2005, only one individual, common to both methods, was assigned as a probable trans-basin migrant. In 2006, 13 individuals were identified as migrants by both methods. Of these, four individuals from the North Sea samples were estimated to have a Irish Sea provenance. No individuals from the Irish Sea were considered as migrants by both methods, and one individual from the Atlantic was considered an Irish Sea immigrant. In 2007, only two out of the nine migrants detected in the North Sea came from the Irish Sea, while four migrants into the Irish Sea had an assigned provenance in the North Sea. In 2008, most of the detected migrants found in the North Sea were assigned to other North Sea samples (23 out 27) and only three were considered by both methods as Irish Sea immigrants; On the other hand, half of the migrants detected by both methods in the Irish Sea came from the North Sea (6 out of 11). Across the four years, the number of fish of hypothetical trans-basin origin was low; resulting in less than 1% of the fish collected in the North Sea having an Irish Sea origin, and around 2% of the fish collected in the Irish Sea were estimated to be born in the North Sea. When the individuals in the uncertain 2005 sample, OfF05, were compared to the other available samples in 2005, 10 out of 21 individuals were allocated to InC05 (Table 7.5). For another six individuals, InC05 was the second best fit, suggesting that the 16 individuals of the original sample were unlikely to be caught at Off Flamborough, and most likely represented fish from the Irish Sea. The last

five individuals, those requested at a later date and labelled as OfF05x, showed highest affinity with NeD05 and probably do represent real off Flamborough dab.

Table 7.4: Migrant detection

Estimated number of migrants as assessed by GENECLASS2 with two different methods. Allele: allele frequency method (Paetkau et al, 1995); and Bayes: Bayesian (Rannala & Mountain, 1997). Common: are the number of individuals identified as migrants by both methods. Only individuals with p <0.01 of belonging to where they were caught are reported. Total = total number of migrants identified by GeneClass2. NS= North Sea; IS= Irish Sea; AT= Atlantic Coast of Ireland. XX<-XX: The first basin indicates where the individual was caught, the basin after (<-) indicates the estimated provenance. Mismatch= the number of individuals for which both methods estimated different origin basins. Ind.= number of individuals included in the analysis (total), number of individuals from the NS (NS<-NS), from the IS (IS<-IS) and from the AT (AT-AT). Pecentages indicate the percentage of fish from one basin that are assumed to have originated in another (i.e NS<-IS: % of fish fished in the NS with an IS origin).

	2005				2006				2007				2008			
	All.	Bay.	Com.	Ind.	All.	Bay.	Com.	Ind.	All.	Bay.	Com.	Ind.	All.	Bay.	Com.	Ind.
Total	4	3	1	236	23	25	13	960	15	14	12	861	39	41	34	614
NS <- NS	1	1	0	139	5	6	1	389	7	6	5	543	23	22	19	466
NS <- IS	0	1	0	0%	6	7	4	1.03%	2	3	2	0.37%	4	6	3	0.64%
NS <- AT	-	-			0	3	0	0.00%	-	-	-		-	-	-	
IS <- IS	2	0	0	97	6	6	3	460	0	1	0	318	5	6	5	148
IS <- NS	1	1	1	1%	3	1	0	0.00%	6	4	4	1.26%	7	7	6	4.05%
IS <- AT	-	-			2	1	0	0.00%	-	-	-		-	-	-	
AT<-AT	-	-			0	0	0	111	-	-	-		-	-	-	
AT <- IS	-	-			3	2	1	0.90%	-	-	-		-	-	-	
AT <- NS	-	-			0	2	0	0.00%	-	-	-		-	-	-	
Mismatch							4				1				1	

Table 7.5: Assigment of OfF05 individuals

Individuals in the OfF05 sample were assigned to the other samples available in 2005 (NeD05, AmB05, LyB05, InC05, StB05). In green are the highest probability value, in yellow the second highest probability. At the bottom of the individuals list (in blue) are the five individuals received separately.

Ind.	NeD05	Amb05	LyB05	InC05	StB05
OfF05-013	0.758	0.446	0.341	0.68	0.155
OfF05-014	0.213	0.198	0.133	0.619	0.127
OfF05-016	0.302	0.085	0.202	0.281	0.047
OfF05-017	0.03	0.063	0.031	0.144	0.009
OfF05-018	0.284	0.272	0.484	0.848	0.172
OfF05-019	0.256	0.161	0.154	0.404	0.03
OfF05-021	0.004	0.002	0.011	0.08	0.013
OfF05-023	0.144	0.061	0.14	0.18	0.155
OfF05-024	0.382	0.422	0.663	0.737	0.146
OfF05-025	0.5	0.595	0.435	0.744	0.263
OfF05-026	0.134	0.22	0.078	0.432	0.45
OfF05-027	0.902	0.732	0.713	0.96	0.976
OfF05-028	0.831	0.405	0.415	0.656	0.113
OfF05-033	0.31	0.286	0.53	0.833	0.068
OfF05-040	0.314	0.096	0.12	0.288	0.047
OfF05-042	0.036	0.137	0.056	0.218	0.112
OfF0x-012	0.576	0.553	0.452	0.799	0.329
OfF0x-013	0.633	0.401	0.078	0.332	0.126
OfF0x-014	0.645	0.362	0.412	0.381	0.065
OfF0x-015	0.272	0.11	0.011	0.019	0.007
OfF0x-051	0.002	0.003	0.002	0.001	0

The population admixture proportions (m_Y) of dab living in LyB and InC are reported in Table 7. 6. The standard deviations were relatively large but general patterns can still be drawn from the results. The North Sea influence on LyB is higher than in InC, and increases over the years, while InC seems to be dominated by an Irish Sea component, albeit there is clear influence of North Sea genotypes in 2008. Unfortunately the results of the individual admixture proportion analysis (*h*) revealed that the genetic difference between North Sea and Irish Sea was insufficient to confidently assign individuals to either basins or hybrid status as the confidence intervals included both basins for many individuals (data not shown).

Location	Year	2005	2006	2007	2008
LyB	NS (m_Y)	0.692	0.535	0.818	0.957
	s.d	0.306	0.191	0.194	0.200
	IS (m_Y)	0.308	0.465	0.182	0.043
	s.d	0.306	0.191	0.194	0.200
InC	NS (m_Y)	0.193	-0.054	-0.060	0.447
	s.d	0.161	0.210	0.156	0.252
	IS (m_Y)	0.807	1.054	1.060	0.553
	s.d	0.161	0.210	0.156	0.252
InF	NS (m_Y)			0.974	
	s.d			0.151	
	IS (m_Y)			0.026	
	s.d			0.151	
RwB	NS (m_Y)			-0.071	
	s.d			0.142	
	IS (m_Y)			1.071	
	s.d			0.142	

Table 7. 6: Population admixture proportion (m_Y)

Admixture proportions (***m_Y***) in dab from LyB and InC samples across the years as calculated in ADMIX 2.0. **NS**: admixture proportion corresponding to North Sea; **IS**: admixture proportion corresponding to Irish Sea. **s.d:** Standard deviations around the means based on 10,000 bootstraps resampling over loci.

7.4.4 Temporal Analysis:

The AMOVA suggested that overall no variation in genetic diversity distribution was explained by changes across years (φ_{SY} =0.000; φ'_{SY}=0.000).

All pairwise Θ_{WC} and Θ'_{WC} values for samples from the same location across years were low (average Θ_{WC}=0.001, average Θ'_{WC}=0.004; Table 7.7) and only five comparisons out of 42, had probabilities below 0.05: StB05 with StB07, RyE06 with RyE07, AmB06 with AmB08, and NeD08 with NeD06 and NeD07. None remained significant after Bonferroni correction, suggesting temporal stability of most locations. Correcting for heterozygosity did not increase the value of differentiation substantially for most comparisons, however, some higher values were noteworthy: StB05 exhibited much higher differentiation with all other StB samples than any other within location comparisons, and InC samples tended to be maximally differentiated to the previous and next year samples.

The patterns of genetic distances (Θ'_{WC}) between samples each year were significantly correlated ($p<0.05$) to each other, except for 2005 and 2006 ($r=0.303$, $p=157$) (Table 7.8), indicating that the general structure between samples was temporally stable. Nevertheless, there were some interesting within location trends. Two locations had significant positive relationships between genetic distance and time: LyB and LiV, suggesting a slight change in genetic composition over time. Another sample had a significant negative relationship between genetic distance and time: InF (Table 7.8). The rest of the locations did not show any significantly trend in genetic distance (Θ'_{WC}) over time.

Table 7.7: Pairwise differentiation across years within location

Pairwise Θ_{WC} (above diagonal) and Θ'_{WC} (below diagonal) between samples from the same location across years. Values with associated probabilities below 0.05 are denoted with *. No values were significant after Bonferroni correction (0.05/6=0.008; 0.05/4=0.013)

	Ned05	NeD06	NeD07	NeD08			LyB05	LyB06	LyB07	LyB08
NeD05		0.000	0.000	0.001		**LyB05**		0.000	0.001	0.003
NeD06	0.003		0.000	0.002	*	**LyB06**	0.004		-0.002	0.001
NeD07	0.000	0.004		0.002	*	**LyB07**	0.001	-0.013		-0.001
NeD08	0.005	0.019	0.009			**LyB08**	0.016	0.011	-0.005	
	Amb05	Amb06	Amb07	Amb08			InC05	InC06	InC07	InC08
Amb05		0.001	-0.002	-0.001		**InC05**		0.002	0.001	0.000
Amb06	0.009		-0.001	0.002	*	**InC06**	0.022		0.002	0.001
Amb07	-0.015	-0.008		0.000		**InC07**	0.008	0.022		0.002
Amb08	-0.010	0.007	-0.001			**InC08**	0.008	0.009	0.004	
		Off06	OfF07	OfF08				RwB06	RwB07	
Off06			0.000	-0.001		**RwB06**			0.002	
OfF07		-0.003		0.000		**RwB07**		0.010		
OfF08		-0.018	-0.015					Liv06	Liv07	Liv08
		InF06	InF07	InF08		**Liv06**			0.000	0.001
InF06			0.000	-0.001		**Liv07**		0.002		0.000
InF07		-0.001		-0.001		**Liv08**		0.006	-0.006	
InF08		-0.013	-0.007				StB05	StB06	StB07	StB08
		Rye06	Rye07	Rye08		**StB05**		0.002	0.006 *	0.002
Rye06			0.001 *	0.000		**StB06**	0.013		0.000	0.001
Rye07		0.003		0.000		**StB07**	0.039	0.000		0.002
Rye08		0.001	-0.001			**StB08**	0.024	0.006	0.010	

Section A

	r	p (+ve)	p (-ve)
Ned	0.177	0.332	0.668
Amb	-0.195	0.740	0.260
Off	-1.000	0.667	0.333
Inf	-0.500	1.000	**0.000**
Rye	-0.500	0.335	0.665
LyB	0.922	**0.000**	1.000
InC	-0.989	0.917	0.083
Liv	1.000	**0.000**	1.000
StB	0.342	0.168	0.832

Section B

	2006		2007		2008	
	r	p	r	p	r	p
2005	0.30277	0.157	0.62234	0.015	0.496431	0.034
2006			0.381837	0.026	0.862177	0.003
2007					0.820781	0.002

Table 7.8: Mantel test of correlation of genetic and temporal distance

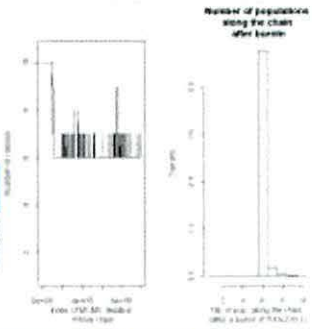
Temporal Mantel Tests: Mantel Test of correlation of genetic distance and temporal distance (years) between samples within locations (**Section A**) and of sample pair-wise genetic distance matrices between two years (2005-2006, 2005-2007, 2005-2008, 2006-2007, 2006-2008, 2007-2008 (**Section B**). **r** = correlation values ; **p** = significance assessed by comparing the actual r with r-scores obtained by randomizing rows and colums of the second matrix.

The temporal analysis in GENELAND suggested six clusters consistently with a minimum probability of 0.7 and a maximum of 0.98, even when populations were ordered randomly (data not shown). The posterior probability maps (Figure 7.10) suggested temporal stability for most locations in the Irish Sea and English Channel (samples for all years of a location fell within the same cluster). Cluster 1 contained those samples from Atlantic (NoI06 and WoI06). The North Sea was mainly placed within Cluster 2, together with RyE and LyB samples from 2006 onwards. Some influence of the Cluster 2 was also evident within the Irish Sea. However, most of the Irish Sea fell strongly within the Cluster 3, except for LiV08, StB08 and StB05 which formed a different cluster (cluster 5). Cluster 4 contained those samples from InC and LyB in 2005. Finally, Cluster 6 was composed of OfF06 and InF06.

The sPCA of the artificial geographic and temporal matrix provided similar results to those from GENELAND, albeit dividing the samples into two major clusters (Figure 7.11). The eigenvalues suggested the existence of only one global component, and very little local structuring. The first component of the sPCA was consistent with a temporal stability of the North Sea-Irish Sea dichotomy. It also revealed that within the North Sea, the coastal samples (AmB, OfF, and InF) were more strongly differentiated than NeD or RyE. In concordance with the temporal Geneland analysis, early North Sea samples (2005 and 2006) also seem less differentiated from Irish Sea than later samples (2007 and 2008). The Atlantic samples (NoI and WoI) were less differentiated from North Sea than the rest of Irish Sea samples. Probably, the most interesting feature of the analysis is the progressive transformation of the LyB samples, which seem to shift from undifferentiated in 2005 to heavily influenced by North Sea in 2008.

Genetime

NoI	StB	Liv	DuB	RwB	WoI	InC	SCB	CeS	LyB	Rye	InF	OfF	NeD	Amb
9.85	8.02	7.39	7.17	6.98	6.4	5.9	5.7	4.48	0	-2.9	-6	-7.3	-8.2	-9.04



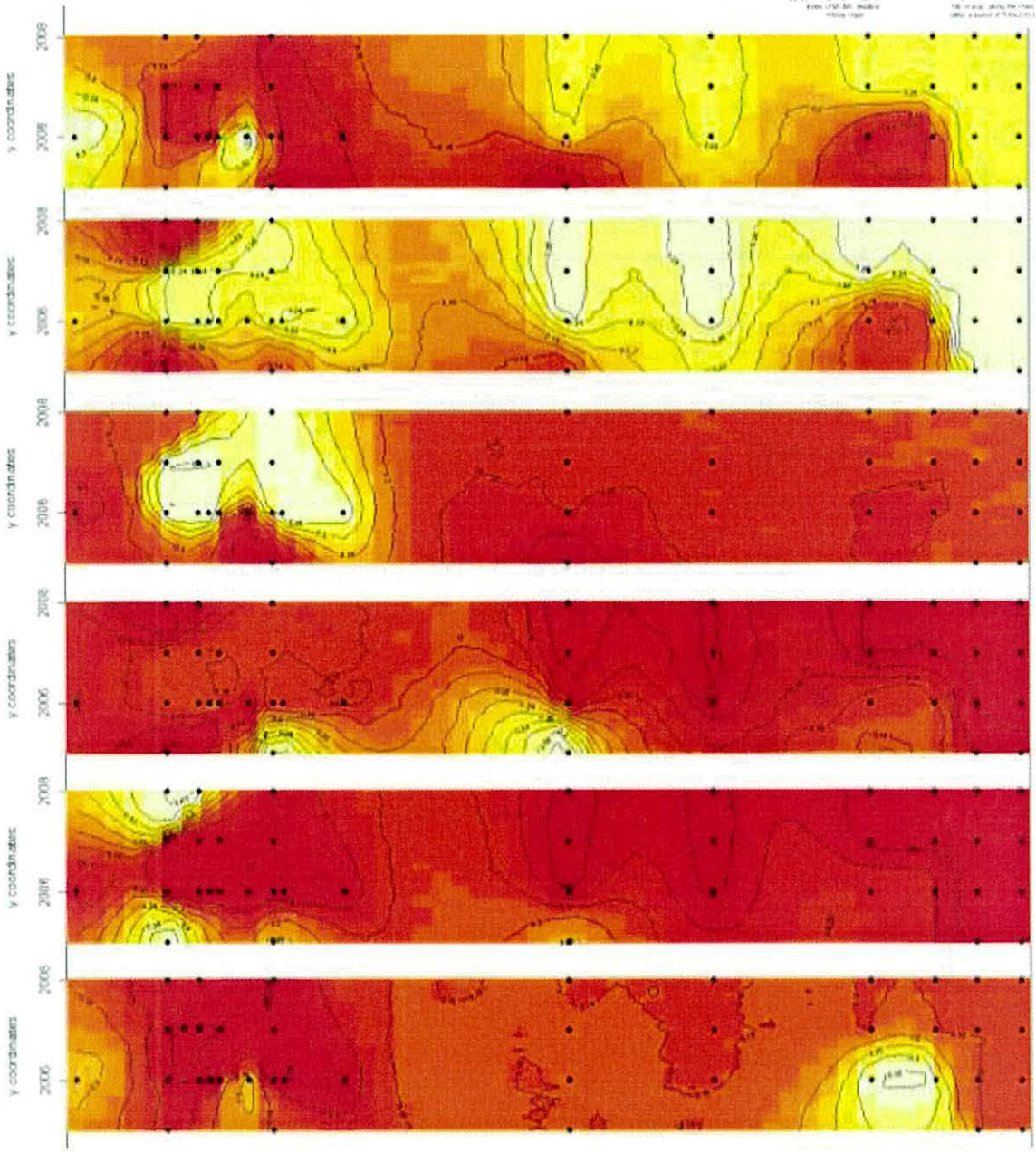


Figure 7.10: Temporal analysis in GENELAND (GENETIME): (Previous page)

GENELAND plot of the artificial geographical and temporal matrix: Location names are listed (dark blue box) from the Atlantic and Irish Sea (left) to North Sea (right). Geographical distances (km) from a centre location (LyB) divided by 100 are listed below (light blue box). GENELAND identified six clusters with very high probability (top left diagram). Therefore, six different membership plots are ilustrated, one for each identified cluster. In each plot the abscissa (x axis) are sampling locations ordered from left to right, while the ordinate (y axis) are years (2005-2008). Each sample is represented by a black dot, and the probability of belonging to a cluster by the intensity of yellow/white. White/yellow indicates high probability of belonging to a cluster, while red indicates low probability of belonging to that cluster.

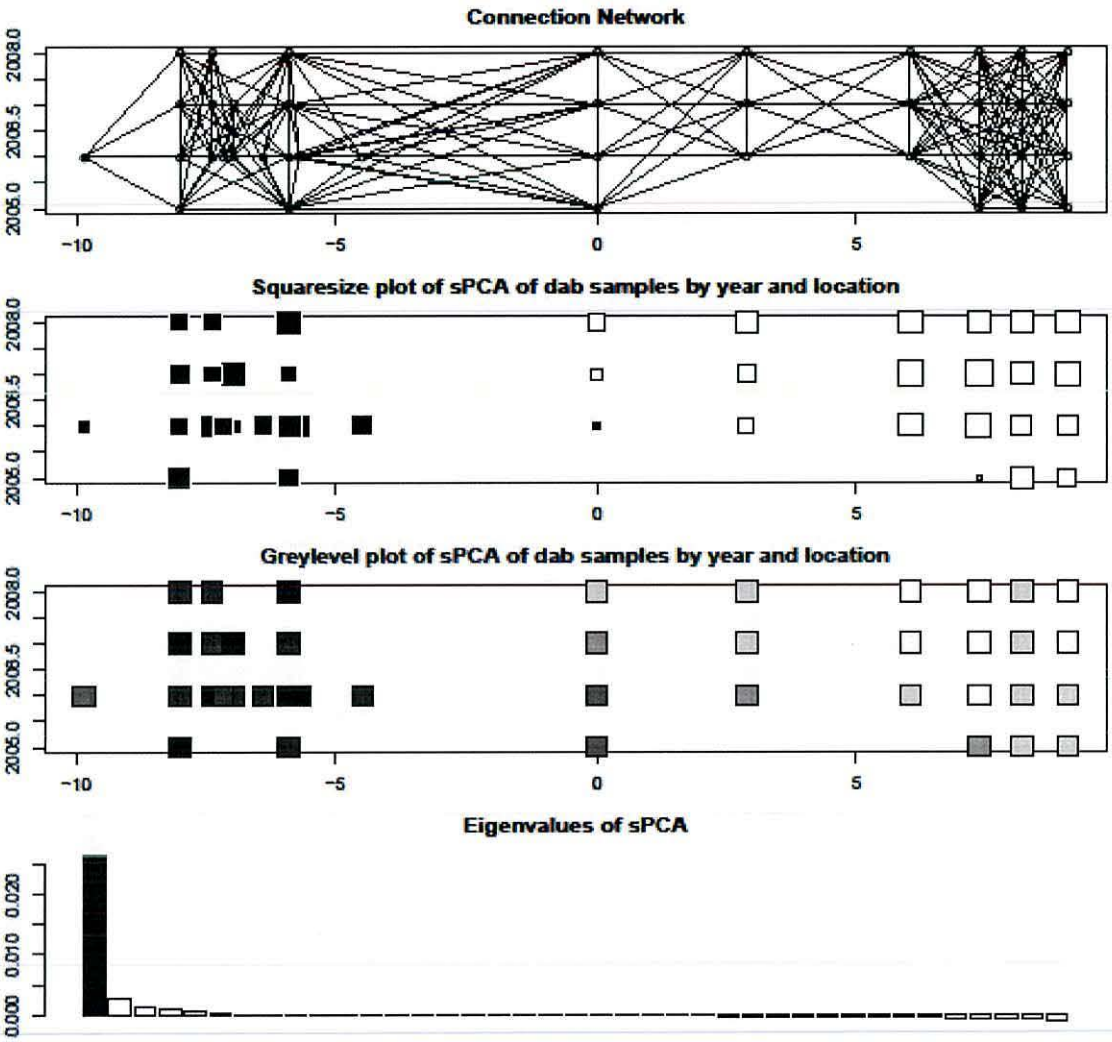


Figure 7.11: Temporal Analysis (sPCA):

A spatial principal component analysis of all dab samples sorted by location (x axis) and year (y axis). Atlantic and Irish Sea samples to the left, English Channel in the middle and North Sea to the right. The order of the locations from left to right is : NoI, StB, LiV, DuB, RwB, Woi, InC, ScB, CeS, LyB(=0), RyE, InF, OfF, NeD, AmB). First plot: the connection network was designed to allow links between all neighbouring samples across all years; Second plot: the size of the square represents absolute values of the first global component, large white squares are well differentiated from large black squares, while small squares are undifferentiated from either. Third plot: as the previous one, but a grey scale is used to indicate undifferentiated samples. Fourth plot: Display of the eigenvalues of the sPCA, indicating the existence of only one major global component.

7.4.5 Age and Sex effects on genetic differentiation:

The age of dab in NeD07 and LiV07 ranged from 1 to 8 years. LiV07 had at least 20 fish in age year classes 1-6, while NeD07 exhibited a bimodal distribution where only age classes 1, 2, 4, and 5 had at least 20 individuals. RyE07 was dominated by fish aged 1 or 2, with very few fish aged 3 or 4. Θ_{WC} values (Table 7.9) between age classes within samples were very low (negative to 0.0014) and non-significant, suggesting genetic homogeneity between year cohorts. Θ_{WC} values between age classes among samples were larger (up to 0.012) and significant in several of comparisons. The ratio of males and females varied widely between samples (10-90%), but were generally female dominated. The proportion of males to females was 50% at age one in NeD07 and LiV07 and decreased to about 20% at age 5. In NeD07 the proportion of males reached 50% again for ages 6 and 7, and in RyE07 the older age classes (3 and 4 year old) were dominated by males (up to 80%). However, the small sample sizes of these older age classes restrict interpretation. The sex-specific GENELAND analysis in 2006 suggested similar groupings for both sexes: females were divided into three groups: coastal North Sea (OfF06 and InF06, and partly AmB06), a second large group containing all of the Irish Sea plus the rest of the North Sea (NeD06, RyE06 and partly AmB06), and a third group with the Atlantic and CeS06 samples. The males were subdivided in four clusters: a first cluster with InF06, a second with the rest of the North Sea, a third conformed of the whole Irish Sea plus CeS06, and finally a fourth clusters with the Atlantic samples. GENELAND clusters for males and females coincided in 2007, with a boundary between InC07 and LyB07. Females in 2008 clustered in two populations, one cluster with LiV08 and StB08 and the rest of the sites in another cluster, suggesting the three individual females of InC08 were genetically more similar to North Sea. While males in 2008 on the other hand separated into three clusters, one with LiV08 and StB08, a second cluster with InC08 (where 27/30 individuals were males) and LyB08, and a third with all the other locations. (Appendix: Table A.1)

7.4.6 Past demographic events:

Most loci in most samples had negative *k* values (average of 13.6 negative loci for all loci, and 11.7 when loci with null alleles were removed), which resulted in significant multilocus *k* for most samples (35 out of 40) (Table 7.10). 17 samples remained significant after Bonferroni correction. No *g* value for any sample was below the cut-off value reported by Reich *et al.* (1999) for the number of individuals and loci studied here. The BOTTLENECK analysis indicated that most samples for all years were significant under the SMM for all three analyses (Sign Test, Standard differences test, and the Wilcoxon test) (Appendix: Table A.2).

Table 7.9: Cohort genetic differentiation

Pairwise differentiation (Θ_{wc}) between yearly cohorts within three dab samples of 2007: North East Dogger (Ned); Rye Bay (Rye); and Liverpool Bay (Liv). Significiance values are denoted with stars (~= near significance;*=below 0.05; **=below 0.01; ***=below 0.001), values significant after Bonferroni correction are in **bold**.

	Ned-1 (69)	Ned-2 (21)	Ned-4 (25)	Ned-5 (32)	Rye-1 (37)	Rye-2 (58)	Liv-1 (42)	Liv-2 (35)	Liv-3 (24)	Liv-4 (28)	Liv-5 (22)	Liv-6 (24)
Ned-1		0.000	0.001	0.000	0.002	0.002 *	**0.004** **	0.004 *	0.006 *	0.004 *	**0.006** **	**0.007** **
Ned-2			0.005	0.001	0.000	0.002	0.005 ~	0.006 *	**0.010** **	0.004	0.003	0.009 *
Ned-4				0.000	0.003	0.000	0.000	0.003	0.002	0.002	0.001	0.006 *
Ned-5					0.002	0.000	0.003	0.004 ~	0.007 *	-0.001	0.008 *	0.004
Rye-1						0.000	0.001	0.001	0.005 ~	0.002	0.002	0.006 *
Rye-2							0.001	0.000	0.004 *	0.001	0.001	0.003
Liv-1								-0.003	-0.002	-0.003	-0.002	0.000
Liv-2									-0.003	0.000	-0.002	0.000
Liv-3										0.002	-0.003	-0.001
Liv-4											-0.004	-0.004
Liv-5												-0.002

Table 7.10: Test of population expansion (*k* and *g*)

Test of population expansion in dab samples from 2005 to 2008 will all loci (16) or all without null alleles (NN). Loc -ve= number of loci with negative *K* values out of 16 (out of 14 for NN). p = multilocus associated probability of *k*, non significant values (p>0.05) highlighted in red, values significant after Bonferroni correction (0.05/40=0.0012) are in bold. *g*= value of the g statistic (all in *italics*), none were below the significance threshold provided in Reich *et al.* 1999.

			NeD	Amb	Off	Inf	Rye	LyB	Ces	Scb	Inc	RwB	Liv	StB	Dub	Noi	Woi
2005	All	Loc -ve	15	11	15			14			13			14			
		p (*k*)	**0.000**	0.085	**0.000**			0.001			0.008			0.001			
		g	*0.907*	*0.765*	*0.939*			*0.987*			*0.817*			*0.952*			
	NN	Loc -ve	13	9	14			12			11			13			
		p (*k*)	**0.001**	0.180	**0.000**			0.005			0.022			**0.001**			
		g	*0.990*	*0.922*	*0.860*			*0.865*			*0.812*			*1.015*			
2006	All	Loc -ve	14	11	13	15	15	13	12	15	15	11	12	13	9	14	12
		p (*k*)	**0.001**	0.085	0.008	**0.000**	**0.000**	0.008	0.029	**0.000**	**0.000**	0.085	0.029	0.008	0.355	0.001	0.029
		g	*0.881*	*0.982*	*0.879*	*0.816*	*1.050*	*0.756*	*0.774*	*0.921*	*1.003*	*0.948*	*0.879*	*1.064*	*1.287*	*0.843*	*0.799*
	NN	Loc -ve	12	9	11	13	13	11	10	13	13	10	11	12	8	12	10
		p (*k*)	0.005	0.180	0.022	**0.001**	**0.001**	0.022	0.073	**0.001**	**0.001**	0.073	0.022	0.005	0.352	0.005	0.073
		g	*0.958*	*0.958*	*0.991*	*0.948*	*1.075*	*0.880*	*0.874*	*0.913*	*0.993*	*0.825*	*0.710*	*1.005*	*0.830*	*0.965*	*0.956*
2007	All	Loc -ve	14	14	14	16	15	15			13	13	13	13			
		p (*k*)	0.001	0.001	0.001	**0.000**	**0.000**	**0.000**			0.008	0.008	0.008	0.008			
		g	*0.908*	*0.827*	*0.890*	*0.878*	*0.945*	*0.912*			*0.886*	*0.937*	*0.974*	*1.095*			
	NN	Loc -ve	12	12	12	14	13	13			11	11	11	11			
		p (*k*)	0.005	0.005	0.005	**0.000**	**0.001**	**0.001**			0.022	0.022	0.022	0.022			
		g	*1.022*	*0.981*	*1.032*	*0.869*	*1.019*	*0.979*			*0.885*	*0.893*	*0.893*	*1.020*			
2008	All	Loc -ve	13	14	15	15	16	13			15	13		11			
		p (*k*)	0.008	0.001	**0.000**	**0.000**	**0.000**	0.008			**0.000**	0.008		0.085			
		g	*0.870*	*0.817*	*0.812*	*0.859*	*0.876*	*1.086*			*2.273*	*0.850*		*0.776*			
	NN	Loc -ve	11	12	13	13	14	11			13	11		9			
		p (*k*)	0.022	0.005	**0.001**	**0.001**	**0.000**	0.022			**0.001**	0.022		0.180			
		g	*1.109*	*0.888*	*0.956*	*0.887*	*0.917*	*0.968*			*0.946*	*0.978*		*0.945*			

7.5 Discussion:

7.5.1 Dab population structure:

The marine environment has often been considered as a rather homogenous realm where ocean circulation and possibilities of individual movement would hinder the development of biological structure (Ward *et al.*, 1994; Waples, 1998). However, instances of significant genetic structure have now been found in a plethora of marine fish species (Galleguillos & Ward, 1982; Pogson *et al.*, 1995; Magoulas *et al.*, 1996; Lundy *et al.*, 1999; Ruzzante *et al.*, 1999; Mattiangeli *et al.*, 2000; van Herwerden *et al.*, 2003; O'Reilly *et al.*, 2004; Mariani *et al.*, 2005; Papetti *et al.*, 2009). Genetic structure in marine species can sometimes even be detected at local geographical scales. Normally these cases are linked to known environmental gradients (Andersson *et al.*, 1981; Bembo *et al.*, 1996; Nielsen *et al.*, 2004; Hemmer-Hansen *et al.*, 2007b; Limborg *et al.*, 2009), but not always (Hutchinson *et al.*, 2001; Bernal-Ramírez *et al.*, 2003; Knutsen *et al.*, 2003; Jorde *et al.*, 2007; Guinand *et al.*, 2008).

Many studies have evaluated population genetic structure in flatfish in the North East Atlantic. Examples include turbot, *Psetta maxima,* (Bouza *et al.*, 1997; Nielsen *et al.*, 2004; Florin & Höglund, 2007), brill, *Scophthalmus rhombus,* (Blanquer *et al.*, 1992), megrim, *Lepidorhombus whiffiagonis,* (Garcia-Vazquez *et al.*, 2006), sole, *Solea solea*, (Kotoulas *et al.*, 1995; Guinand *et al.*, 2008), Greenland halibut, *Reinhardtius hippoglossoides,* (Knutsen *et al.*, 2007), and Atlantic halibut *Hippoglossus hippoglossus,* (Foss *et al.*, 1998; Reid *et al.*, 2005). However, few surveys have included samples from both the Irish and North Seas, perhaps symptomatic of the expected lack of structure at such a localised regional level. In those studies which included samples from both the Irish and North Seas, significant genetic structure as revealed here in dab has been found in sole (Exadactylos *et al.*, 1998), but is generally absent from other similar marine flatfish like flounder (Galleguillos & Ward, 1982; Borsa *et al.*, 1997; Hemmer-Hansen *et al.*, 2007b; 2007a) and plaice (Hoarau *et al.*, 2002b; 2004; 2005).

Given that dab is the third most common fish in North Sea (Daan *et al.*, 1990), the relatively recent colonisation by North and Irish Sea biota after the last glacial maximum (Maggs *et al.*, 2008), and potential for egg and larval dispersal (Henderson, 1998), little genetic structuring would be expected. Contrary to such expectations, temporally stable patterns of significant differentiation were observed. Dab around the British Isles exhibits detectable genetic structure where fish from the North Sea were consistently genetically distinct from those in the Irish Sea proper. The differences between North and Irish Seas were evident with all tests

undertaken, and strongly significant. The temporal stability of such structuring indicates that it is not the outcome of the random distribution of genetic diversity, but more likely the result of biological processes shaping such distribution (Waples, 1998; Waples *et al.*, 2008).

A strong isolation by distance pattern was detected at first using the full data set; however, controlling for sea basin reduced the associated probabilities to non-significant in most years. Taken together with the absence of isolation by distance patterns within basin, and the lack of relationship between genetic and geographic distances when classed into within and between basins, is it evident that the distribution of genetic diversity was little influenced by geographic distance. It thus becomes more likely that dab populations are subdivided by the coastal and oceanographic features of the British Isles. Most of the genetic diversity of dab was explained by differences within individual dab (98.4%), which is common in many marine fish (Gyllensten, 1985; Galvin *et al.*, 1995; Giæver & Stien, 1998; Lundy *et al.*, 1999; van Herwerden *et al.*, 2003; Nielsen *et al.*, 2004; Jorde *et al.*, 2007). However, the among-individuals, among-samples and among-basins components of genetic diversity were all highly significant overall and for most years. Once corrected for heterozygosity, AMOVA suggested that sea basin explained around 2% of the genetic variation found in dab, while only 0.4% differentiated samples within basin. The K-means clustering analysis and the sPCA both suggested the existence of two broad groups for all years: one composed of all the North Sea and English Channel samples, and the other comprising fish from the Irish Sea and Atlantic.

Samples collected from the North Sea were generally considered very similar to one another in most of the analyses performed: structuring within the North Sea was much weaker. Nevertheless, the sPCA depicted the NeD and RyE samples as marginally less differentiated from the Irish Sea, than the other coastal North Sea samples (AmB, OfF, and InF), a pattern that was also supported by the temporal sPCA. Such differences were also detected by the 2nd and 3rd order barriers in the Barrier analysis dividing NeD from the rest of North Sea. The NeD samples were also significantly differentiated (although not always after Bonferroni correction) from southern North Sea samples (InF and/or RyE) in the three years examined. Overall, despite the lack of significance in some years, the different test suggests very weak differentiation across the North Sea, although perhaps not temporally stable. Structure at such low differentiation values could be caused by differences in sample size, or slight changes in genetic composition between cohorts.

Within the North Sea, the Off Flamborough 2005 sample was exceptional in that it clustered with the ISS samples in the neighbour-joining trees, k-means clustering analysis, GENELAND, CA and sPCA. However, two issues have been identified which may have resulted in such pattern: first, the sample size (N=21) was the smallest used employed, and may have biased the heterozygosity and allele frequency within the sample, thus making it unreliable; and secondly, although the samples were labelled as belonging to Off Flamborough when received, they were located in the middle of Irish Sea samples, and thus treated with caution ever since. The individual assignment analysis revealed the origin or the disparity; 16 of the samples were assigned to the InC05 population, while the other five, which were shipped separately, clustered with the NeD05 sample, and thus probably represented a fraction of the real OfF05 sample. Emphasis needs to be placed in that OfF05 was the only sample for which there were doubts about the origin of individuals, and the potential problem with the sample was identified from the outset.

The Irish Sea was also quite homogeneous with very few significant genetic differences between samples: InC05 and StB05, and RwB07 with InC07 and LiV07. However, the Irish Sea can be divided into two: The northern part (ISN: RwB, LiV, StB), which consistently clustered together in all analyses; and the samples collected from Cardigan Bay (ISS), which some analyses and years grouped with ISN, while in others it clustered with LyB. Of particular note are the results of the Barrier analysis: most other analysis placed the confines of the Irish Sea dab population south of InC or east of LyB, while Barrier suggested the most supported edge (bootstrap and Loci wise) for the first three years was north of InC. Barrier highlights the sharpest change in genetic distance between two adjacent samples, thus suggesting that gene flow between InC and ISN may be somewhat limited although not resulting in as strong genetic differentiation as between sea basins.

The structure in the English Channel was more complex and varied across years. With the exception of 2006, genetic differentiation between the two English Channel samples was not significant. However, their affinity to other sea basins differed substantially. Rye Bay was found to be more closely related to North Sea than to Irish Sea, and neighbour joining trees, CA, GENELAND, and sPCA always placed it within the North Sea group. On the other hand, Lyme Bay changed genetically over the course of the study. The positioning of InC and LyB samples in either North Sea, Irish Sea or their own group, seems to be a recurrent trend throughout the years and across analysis methods. First, LyB seemed undifferentiated from either North Sea nor Irish Sea as pairwise genetic differentiation values between LyB and

other samples (North Sea and Irish Sea) were smaller and less significant than between other cross-basin comparisons. The sPCA also depicted LyB samples as undifferentiated from either Irish Sea or North Sea. Although the K-means clustering analysis placed LyB within the North Sea cluster, it was always the first sample to be separated from North Sea when the number of clusters was increased. With GENELAND LyB samples clustered with either the North Sea (2006, 2007) or together with InC in a third cluster (2005, 2008). LyB samples played a crucial role in creating the illusion of an isolation by distance pattern in the studied system, as their undifferentiated nature fulfilled the middle link between North Sea and Irish Sea.

In those samples for which age information was available, the proportion of females increased with age. The prevalence of males at younger age classes and subsequent female domination in older age classes has been reported in some studies (Albert *et al.*, 1998; Deniel, 1990). There were differences in the significance of sex-specific patterns of differentiation. For example, females in 2006 displayed considerable levels of differentiation within the North Sea, and those from ISS were in general more differentiated (Θ_{WC}) from North Sea than males from the same area. On the other hand, females from the ISN in 2006 were not statistically different from those in the North Sea. However, no clear pattern emerged throughout the years. Furthermore, although the AMOVA suggested different φ_{ST} values for males and females every year, the standard deviations overlapped between the sexes, thus suggesting that the differences the sexes are not statistically significant and may be random.

As mentioned earlier, no significant genetic structure has been found among other flatfish at the regional scale covered here (Galleguillos & Ward, 1982; Borsa *et al.*, 1997; Hoarau *et al.*, 2002b; 2004; 2005; Hemmer-Hansen *et al.*, 2007b). These studies have shown significant structuring between the European continental shelf and either Iceland or the Baltic Sea, but have found no significant structure within the continental shelf itself. The reasons for the lack of structure in flounder and plaice may be at the experimental design level: the fine-scale sampling, increased number of loci in this study (16), and temporal replication of the same sampling sites may have enhanced our ability to detect significance of structure; alternatively, dab biology may be less affected by some of the characteristics that typify many other marine fish such as long ranging pelagic eggs and larvae and strong migratory behaviour.

Dab eggs and larvae are ubiquitous throughout the North Sea (Rijnsdorp *et al.*, 1992) and the Irish Sea (Fox *et al.*, 1997). However eggs can hatch quickly (4.5 days at 14°C), and larvae seem capable of controlling, to some extent, their movements (Henderson, 1998; Beggs & Nash, 2007), which together with their less demanding nature regarding depth than other

flatfish larvae (Bolle *et al.*, 1994; Gibson *et al.*, 2002), may effectively reduce the effective dispersal of most individual eggs. Furthermore, Galarza *et al.* (2009) verified that marine currents and sea fronts can represent effective barriers to gene flow between populations, regardless of egg type or pelagic larval duration, and results in detectable genetic differentiation at regional scales. The cyclonic gyre that forms within the Irish sea in spring and summer (Hill *et al.*, 1997) has been suggested as a larvae retention mechanism for Norway lobster, *Nephrops norvegicus,* (Hill *et al.*, 1996), and could have a similar effect on dab eggs and larvae. Furthermore, a strong jet-like westward flow across the St. Georges Channel (between St. Davids Head of Wales and Carnsore Point in Ireland) prevents the blend of Celtic and Irish Seas water masses (Brown *et al.*, 2003), further complicating the export of eggs and larvae between the English Channel and North Sea. On a modelling exercise incorporating oceanographic data of the Irish Sea, simulated eggs and larvae dispersed an average of 80km from point of release, largely remained within 160km of point of release, and very few travelling up to 300km from point of release (van der Molen *et al.*, 2007). These dispersal distances and the oceanographic features of the Irish Sea are in accordance with the structure found here in dab between North Sea and Irish Sea, for which a major genetic boundary exists between InC and LyB dab, and are suggestive of lack of migrants during the early life history stages of dab.

Adult dab are known to engage in seasonal migrations between feeding and spawning grounds, but these are generally thought to be between the coastal and offshore sections of certain areas (i.e from the Wadden Sea to the German Bight) (Rijnsdorp *et al.*, 1992), not between basins. The assessment of the real time migrant analysis suggested little exchange of migrants.

7.5.1.1 Adult migration:

Levels of differentiation as low as φ_{ST} = 0.009 and Θ_{WC} = 0.027 were considered sufficient by Ruzzante *et al.* (2006) to quantify the relative contribution of different spawning subpopulations of Atlantic herring, *Clupea harengus,* to fished feeding aggregations. Similarly, Hauser *et al.* (2006) evaluated that Θ_{WC} as low as 0.02 were sufficient to clearly distinguish between wild steelhead salmon, *Oncorhynchus mykiss,* and their hatchery born counterparts in creeks were the latter had been released. Jorde *et al.* (2007) studied cod collected along a 200 km stretch of Southern Norway coast and was successfully reassigned between 50 to 70% of individuals back into the population of origin even with an average Θ_{WC} as low as 0.0013. Despite being low, the genetic differentiation found in dab between sampling sites was still enough for the detection of probable migrants. Note however, that two key issues hinder the

accuracy of the results: first, not all possible donor populations have been sampled, which in the previous examples (Hauser *et al.*, 2006; Ruzzante *et al.*, 2006) was paramount for the correct assignment of individuals; and second, only migrants with outstanding genotypes could be identified, as given the low differentiation, many combinations of genotypes are probably common in all basins.

Over the course of four years, 21 dab out of 3027 were considered trans-basin migrants by both allele frequency and Bayesian methods. Although the number of migrants was estimated to be low, rare long distance migration rates of 1% per generation as found here would very quickly erode the genetic differentiation found in dab (Slatkin, 1987; Selkoe *et al.*, 2008; Waples *et al.*, 2008). Two potential hypotheses would explain the existence of detectable immigrants: first, the individuals identified as migrants are actually fish locally born but carry a particularly rare combination of alleles at the studied loci; conversely, detected migrants may indeed be true long distance migrants (with a genotypes typical of elsewhere), but these migrants will not carry suites of locally-adapted genes which may prevent them from successfully contributing to the next generation. Such an adaptive scenario would make compatible the occurrence of long ranger migrants together with the existence of genetic differentiation.

7.5.2 Temporal aspects of genetic structure:

The reproducibility of patterns of genetic structure over time strengthens the biological significance of the differences found (Carvalho & Hauser, 1998; Waples, 1998). Some studies have assessed the stability of patterns of genetic structuring over long periods of time, either by sampling the same locations after a number of years or analysing stored biological material (Bernal-Ramírez *et al.*, 2003; Hutchinson *et al.*, 2003; Hoarau *et al.*, 2005; Poulsen *et al.*, 2006); while other studies, like the current one, have aimed at finding how stable the structuring is in consecutive years (Bembo *et al.*, 1996; Mariani *et al.*, 2005; Ruzzante *et al.*, 2006; Hemmer-Hansen *et al.*, 2007a; Pampoulie *et al.*, 2008; Franckowiak *et al.*, 2009; Papetti *et al.*, 2009). Some have found largely stable genetic patterns (Bembo *et al.*, 1996; Bernal-Ramírez *et al.*, 2003; Poulsen *et al.*, 2006; Ruzzante *et al.*, 2006; Hemmer-Hansen *et al.*, 2007a; Pampoulie *et al.*, 2008), while others have reported both long term changes and genetic differences between consecutive years suggesting the structure was transient or was altered by external factors (Hutchinson *et al.*, 2003; Hoarau *et al.*, 2005; Mariani *et al.*, 2005; Franckowiak *et al.*, 2009; Papetti *et al.*, 2009).

The temporal stability across consecutive sampling years of the genetic structure of dab was remarkable. No significant genetic differentiation (Θ_{WC} or Θ'_{WC}) was found at any location between the temporal replicate samples. The differentiation values were very low, even after correction for heterozygosity, highlighting that even highly polymorphic and structured loci (such as *DAC4-64*) had similar allele distribution within location across years. Both the temporal GENELAND and sPCA suggested genetic stability for most locations, indicative that the structure found for dab is likely to be biologically significant rather than the product of random distribution of alleles (Carvalho & Hauser, 1998; Waples, 1998).

Dab targeted in the stations covered by the CSEMP programme seem younger than previously reported in the early 1990's, where the 5 and 6 year old classes accounted for the majority of the dab population in the Eastern North Sea, with very few one year old fish collected (Rijnsdorp *et al.*, 1992). Such changes in age composition of populations could be explained by age segregation, increased mortality, or temporal changes in recruitment numbers. Indeed cohort strength fluctuations have been reported for dab (Henderson, 1998; van der Veer *et al.*, 2000; Beggs & Nash, 2007), which could generate apparent genetic structuring. Nonetheless, the temporal stability was evident even among age cohorts for those samples for which age information was available (NeD07, RyE07 and LiV07), suggesting that the stability of genetic composition was inherent to the local dab population and not a chance effect of sampling the same very successful cohort across the four years. Our results are thus more in line with those obtained among cod cohorts, where differences among cohorts within samples were not significant (Jorde *et al.*, 2007), than those from herring, where a significant effect was detected among year classes within samples (Mariani *et al.*, 2005).

Despite general temporal stability, there were exceptions: The StB05 sample was more differentiated from the rest of StB samples than any other within-location comparison, though, such pattern was most likely due to its smaller sample size (N = 30), especially when incomplete genotypes were removed (N = 18). Without doubt the most interesting temporal change is that found in the western English Channel: LyB samples became increasingly more differentiated over time, a relationship picked up by the Mantel test ($r=0.922$; $p<0.001$) and also displayed by the temporal GENELAND and sPCA. The CA also placed the 2005 LyB sample very close to those from the Irish Sea; while those in 2006 and 2007 were located in between both basins; finally, the 2008 sample was clearly placed in the middle of all the North Sea samples. The effect of North Sea dab moving into LyB was also suggested by the outputs of the Barrier analysis, as the most supported barrier shifted from north of InC for the first three

years to south of InC, between LyB and InC, in 2008. Besides, 2008 was the only year for which all estimators of differentiation agreed on the first barrier, suggesting the strong shift in allele frequencies created by the direct contact between North Sea genotypes and Irish Sea genotypes eclipsed the change of allele frequencies north of InC previously detected in other years. These patterns, together with the increase in North Sea admixture proportion over time are suggestive of an increase of North Sea dab into the western English Channel. Unfortunately the individual admixture analysis could not resolve whether the change was due an increase of pure North Sea dab (immediate recent migrants) or an increase in the North Sea proportion of the genome of all dab (diluted effect of more ancient migrants).

7.5.3 Past demographic events:

The past demographic history of a population is key in understanding the present distribution of genetic diversity. Patterns of recent migration can result in no genetic differentiation (Slatkin, 1987); but so can recent divergence of two populations from a common ancestor. The recent creation of North and Irish Seas (Behre, 2007) have been hypothesised as causes for the presence of only weak genetic structure found many other marine organisms with large population sizes in the North East Atlantic (Gyllensten, 1985); and evidence of expansion signals at both mtDNA and microsatellite data have been detected in many species such as cod, *Gadus morhua* (Pogson *et al.*, 1995; Pampoulie *et al.*, 2008), flounder (Borsa *et al.*, 1997), plaice (Hoarau *et al.*, 2004), common crabs, *Carcinus maenas* (Roman & Palumbi, 2004), gobies, *Pomatoschistus microps* (Gysels *et al.*, 2004), thornback rays, *Raja clavata* (Chevolot *et al.*, 2006), and fucoids, *Fucus serratus* (Hoarau *et al.*, 2007).

Most loci for most samples revealed negative *k* values suggesting alleles for these loci all emerged in a short period of time. The exclusion of those loci with null alleles (*DAC1-35* and *DAC5-70*) only affected the significance class of five of the 40 samples; furthermore, Reich *et al.* (1999) discourage selectively removing loci from analysis, and therefore only the values for the whole locus suite will be discussed. The multilocus associated probabilities for most samples (35 out of 40) were below 0.05, strongly suggesting dab, as a whole, have experienced a dramatic population expansion in the past. Only five samples failed to reach the 0.05 significance level. In the North Sea: Amble both in 2005 and 2006; and three samples in the Irish Sea: RwB and DuB in 2006, and StB in 2008. On the other hand, the distribution of values significant after Bonferroni correction was not random either: InF and RyE were always highly significant; NeD, OfF, and samples in the Cardigan Bay area also seem prone to

significance; LyB was significant in 2007; while no *k* value for Irish Sea (North) samples remained significant after Bonferroni correction.

On the other hand, no *g* value was significant. However, the interlocus *g* test is particularly sensitive to variations in mutation rates between loci, which increase the ratio of *g* (Reich & Goldstein, 1998), and other studies have also reported similar patterns of highly significant *k* values associated with non-significant *g* values (Donnelly *et al.*, 2001; Munkacsi *et al.*, 2008). Unique locus characteristics such as homoplasy, allele size restrictions, and whether the polymorphism at the locus is species specific or was present before the creation of the species under study, would all result in widely divergent estimates of the age of the deepest allele branching effectively blurring the expansion signal as detected by the *g* ratio. As in other studies using *k* and *g* tests (Donnelly *et al.*, 2001; Munkacsi *et al.*, 2008) and given the overwhelming significance of the *k* statistic in most dab samples the population expansion signal is accepted as true. Two possible past events could explain an expansion signal in dab: the colonisation of the North Sea and Irish Sea after the last glaciation (Maggs *et al.*, 2008), or the recent surge in dab numbers reported since the 1970's (Kaiser & Ramsay, 1997). These hypotheses will be discussed below.

The original aim of the BOTTLENECK analysis was to resolve whether those samples not showing an expansion signal (i.e. AmB in 2005 and 2006 and the Irish Sea) had instead recently experienced a recent contraction in population size. The results provided unexpected insights into the past demography of dab. Most samples had H_E outside the expected H_{eq} 95% CI under the SMM, the most conservative mutation model and the one that best applies to microsatellites. However, the Wilcoxon test indicated that the source of the significance was heterozygote deficiency instead of excess, suggestive of population expansion (Luikart & Cornuet, 1998). A more detailed locus specific analysis (Data not shown) revealed some very interesting patterns about the population expansion signal and microsatellite evolution: first of all, four loci (*DAG2-90*, *DAG4-64*, *DAG5-17*, and *DAC1-90*) showed highly significant heterozygosity excess for most samples under the IAM. Such a pattern would indicate that a bottleneck has affected all samples only in the part of the genome where these loci reside, however, it is much more plausible that such a pattern is just indicative that these loci are very unlikely to follow an IAM. When locus-specific values were analysed under the SMM the opposite pattern was revealed, most loci were highly significant in most samples but for heterozygosity deficiency, indicative of population expansion over most of the genome. The four loci aforementioned, together with *DAC5-21*, did not show deviations from the expected

H_{eq} for any sample. Such an apparently paradoxical pattern becomes sensible when the power of the analysis is examined. As reported by Cornuet & Luikart (1996), the power of detecting population fluctuations diminishes faster with time since bottleneck/expansion for highly diverse loci ($H_E < 0.8$; detectable for $0.1x2N_e$ generations) than for loci with average H_E ($0.3 < H_E < 0.8$; detectable for $0.25x2N_e$ generations). All loci which complied with the expected H_{eq} had $H_O > 0.9$. The volatility of demographic change signals in highly diverse loci is explained by Cornuet & Luikart (1996) by their higher mutation rate, and thus a quicker recovery of H_{eq}. In the case of dab, such property implies that the allelic diversity has reached equilibrium with heterozygosity in highly diverse loci but not the rest of the genome. Such a recovery would be impossible from a very recent expansion signal such as that created by the current increase in dab numbers (Kaiser & Ramsay, 1997), thus the alternative explanation, recent colonisation of the Irish Sea and North Sea, will be explored. D'amato (2006) employed BOTTLENECK to evaluate past demographic changes in long-tailed hake, *Macruronus magellanicus*, and also found significant heterozygosity deficit with compared to H_{eq}, which was linked to population expansion as well. In agreement with the dab results, the most diverse microsatellite loci used by D'amato had reduced signals of population expansion.

The population expansion signal found here with two different methods, allele distribution (*k* test) and heterozygosity (BOTTLENECK), is consistent with the recent re-inundation of the North and Irish Seas 10,000 and 8,000 years ago (Behre, 2007) and subsequent colonisation by its current biota (Maggs *et al.*, 2008). If a dab generation time of two years for the males and three years for females is assumed (Bakhsh, 1982; Rijnsdorp *et al.*, 1992), dab must have expanded into the North Sea and Irish Sea about three or four thousand generations ago, which is similar to the number of human generations estimated (4800) since the population expansion detected (60,000 to 100,000 years ago) by *k* and *g* tests in African human populations (Reich *et al.*, 1999). At the height of the last glacial maximum, dab census size must have been much smaller than today as suitable habitat for a shallow demersal species such as dab, would have been limited to the narrow Atlantic coastal shelf off France, Spain, and Morocco (Figure 7.12;Figure 7.13). As the ice sheet retreated, the large shallow expanses of continental shelf around the British Isles must have represented a formidable habitat expansion for dab, as has been shown for many sub-boreal marine organisms (Pogson *et al.*, 1995; Borsa *et al.*, 1997; Hoarau *et al.*, 2004; Gysels *et al.*, 2004; Chevolot *et al.*, 2006; Pampoulie *et al.*, 2008).

Figure 7.12: Map of coastal Europe.

Map of Europe showing the width of the coastal shelf. Reproduced from Google™Earth. Light blue represents the coastal shelf with shallow depths. Darker blue represents the deeper ocean basin. Note the narrowness of the coastal shelf around the Bay of Biscay, the Iberian Peninsula and Morroco. The location of the sampling stations is provided for reference.

Figure 7.13: Map of coastal Europe at the Last Glacial Maxima.

Reconstructed map of Europe when Ice sheets were at their maximum expanse (21,000 years ago). Borders of ice sheets represented in light blue (Maggs *et al.*, 2008); Whether the British and Scandinavian Ice sheet joined is subject to controversy; Postulated North Sea and English Channel shorelines 12,000 years ago (65m below current sea level) (Behre, 2007) are approximately delineated in brown.

Dab has an estimated biomass of two million tonnes in the North Sea (Daan *et al.*, 1990), which together with the average weight of adult dab (ca. 100g; average weight of dab included in the CSEMP programme), translates into 20,000 million individual dab in the southern North Sea alone. Although such estimation is very crude, it still conveys an approximation of the magnitude of the number of dab living around the British Isles. For organisms with such large census population sizes, genetic drift is not likely to play a major role on population differentiation, and thus differentiation at neutral markers between isolated dab groups is expected to emerge very slowly.

7.5.4 Conclusion and brief implications for biomonitoring programmes:

Overall, the results of current study reveal the existence of significant genetic structure of dab at the regional scale covered here, in which dab from the different sea basins were genetically distinguishable. Several non-mutually exclusive hypothesis can lead to structure in the marine environment: reduced migration combined with genetic drift (Borsa *et al.*, 1997), fidelity to spawning site (Thorrold *et al.*, 2001; Vitale *et al.*, 2008), and local adaptation (Carvalho, 1993; Conover *et al.*, 2006; Hemmer-Hansen *et al.*, 2007a). The small *Fst* values found here may not necessarily imply high levels of gene flow between populations (Hauser & Carvalho, 2008). On the contrary, the existence of small *Fst* values with neutral markers despite such large populations, and recent colonisation history, may suggest the existence of biologically significant genetic differences at the adaptive level and the existence of locally adapted populations. Several authors have previously studied different dab populations across its distribution range, and have reported differences among them in morphological (e.g. number of vertebrae, spines and rays) and life history traits (e.g. growth rate, maximum size and age at maturity) (Bakhsh, 1982; Deniel, 1990; Rijnsdorp *et al.*, 1992; Henderson, 1998). The results from the current study raise the possibility that such differences may not just represent phenotypic plasticity, but could indeed be genetically-based, indicating a higher than expected degree of isolation among populations. Therefore, their mechanisms of response to pollution and other environmental stresses might vary between basins. Evidence of such patterns has already started to emerge: blood plasma proteome of dab not only reveals their liver tumour status, but also their provenance: Irish Sea or North Sea (Ward *et al.*, 2006); larger sex biases have been reported in expression of hepatic ethoxyresorufin-O-deethylase (EROD) in North Sea fish than in Irish Sea fish (Kirby, 2003); and dab from the North Sea and Irish Sea seem to have a two year gap between the earliest onsets of the cancerigenous process in each location (Grant Stentiford, CEFAS, personal communication). Furthermore, the disease profiles of dab seem to be temporally stable (Stentiford *et al.*,

2009), reinforcing the idea of locally stable dab populations. Given the patterns of genetic differentiation found in dab around the British Isles, it is recommended that biomonitoring programmes take such structuring into account: fish from target polluted sites should be compared to fish from reference sites *within* the same subpopulation boundaries. In doing so, managers would avoid confounding variance due to subpopulation-specific inherent likelihood of developing disease and that truly caused by pollutants. Additionally, subpopulation-specific disease-frequency baselines can now be established with evolutionary meaningful geographical boundaries, allowing detection of finer changes in location-specific disease prevalence.The genetic changes reported at the western English Channel may command special treatment in biomonitoring analysis, as the susceptibility of the immigrant North Sea fraction of the population may be different from the originally found in LyB.

Overall, the microsatellites developed for dab (Tysklind *et al.*, 2009b) have proved invaluable in detecting subtle but highly significant and temporally stable genetic structure around the British Isles. The acknowledgement of such structure by the biomonitoring community will contribute to a better understanding of the processes undertaken by dab in their quest for adaptation and survival in an increasingly human-altered environment.

7.6 Summary:

1) Over 3,000 dab collected from up to 15 stations over the course of four years were genotyped at 16 microsatellite loci.

2) Significant structuring of the genetic diversity of dab was found, suggesting that dab around the UK form at least two distinct subpopulations.

3) The genetic structuring was found to be generally stable over the course of four years, indicating the structuring is due to biological reasons and not a sampling artefact.

4) The main subpopulations found, North Sea and Irish Sea, correspond to differences found in other traits of dab biology (e.g. physiological and life-history-traits).

5) Biomonitoring programmes using dab around the British Isles should be aware of the evolutionary isolation of dab subpopulations and analyse the biomonitoring data accordingly (i.e. only comparing within subpopulation target and reference sites).

Chapter 8: Integration of population genetic data into marine biomonitoring: a case study using dab, *Limanda limanda*.

8.1 Abstract:

The dab, *Limanda limanda,* is one of the key biomonitoring species in the UK and is extensively studied for biological responses of exposure to anthropogenic pollutants. Using a mobile species, however, to evaluate localized pollution burdens can be problematic, as sampling location might not represent the true life-long exposure record. As revealed by genetic markers dab are not a single homogenous entity throughout the sampling range. Here the patterns of genetic connectivity and disease incidence are compared and a disparity between the strongest barrier to gene-flow and the sharpest change in biomarker profiles is found. There was no obvious relationship between two measures of inbreeding and disease records, indicating that mating between close relatives is not the cause for individual difference in biomarkers. Furthermore, there was no evidence for increased relatedness among individuals afflicted with liver nodules, suggesting that carcinogenic processes are not restricted to a particular genetic section of the population. Genetic differentiation among samples is exploited to study the provenance of fish displaying liver nodules. Most diseased fish can be assumed to belong to where they were collected; however for some samples with abnormally high frequency of liver nodules, like St. Bees in 2007, the incidence can be explained by recent immigration from other locations. Genotyping of assessed individuals provided important information not available by other means and the incorporation of population genetic data is encouraged for programmes studying mobile species.

8.2 Introduction:

Biomonitoring programmes aim to assess the effects of anthropogenic pollutants in the natural environment. They use bioindicator species for which biomarkers (parasite loads, diseases and other conditions) associated with pollution exposure are recorded. The incidence of each biomarker can then be studied spatially and temporally to understand the biological implications of chemical compounds derived from human activities (Phillips & Segar, 1986). In aquatic ecosystems fish are commonly used to evaluate the effects of

industrial and domestic effluents on rivers, lakes, estuaries and coastal areas (van der Oost *et al.*, 2003), and biomonitoring exercises employing fish abound (MAFF, 1987; Förlin & Celander, 1993; Schmitt *et al.*, 1993; Teh *et al.*, 1997; Reichert *et al.*, 1998; Feist *et al.*, 2004; Breine *et al.*, 2007; Schlacher *et al.*, 2007; Leonardi *et al.*, 2009).

Dab (*Limanda limanda*) are one of the key biomonitoring species in marine environmental quality assessment in the North East Atlantic, and monitoring programmes, both at European (OSPAR, 2009) and UK levels (CEFAS, 2003c; 2005; Feist *et al.*, 2008) have been established. As part of the Clean Seas Environmental Monitoring Programme (CSEMP) a sampling cruise is undertaken every year to assess the health status of dab around the UK. On a standard CSEMP station, disease profiles and parasite load of individual fish are recorded. Upon collection dab are immediately placed in flow-through tanks with aerated sea water, and individual fish evaluated in one of two ways in each sampling station: first, a set of 50 fish, described here as "comprehensive", are thoroughly examined for external diseases and parasite load (Bucke *et al.*, 1996; Broeg *et al.*, 1999; Schmidt *et al.*, 2003), the liver is sampled for histopathology (Feist *et al.*, 2004) and the gonad analysed for evidence of endocrine disruption (Scott *et al.*, 2007). Additionally, bioindicators of exposure to organic contaminants such EROD activity and bile metabolites are examined (Kirby, 2003; Neall *et al.*, 2003), DNA is assessed for evidence of disruption (Everaarts, 1995; Lyons *et al.*, 2000; CEFAS, 2003b), and blood plasma is collected for proteomic analysis (Ward *et al.*, 2006), collectively yielding a very complete individual disease record. Complementarily, larger numbers of fish (up to 300 in some stations), and denoted as "externals only", are screened for external diseases and grossly visible liver pathologies. The combination of both methodologies generates detailed information of disease processes (comprehensive) together with a more accurate representation of disease frequency at each location (externals).

Dab display variations in the incidence of diseases and other biomarkers of pollutant exposure among sampling locations (Lyons *et al.*, 2006; Ward *et al.*, 2006; Stentiford *et al.*, 2009) which are then reported and interpreted in evaluations of ecosystem health (MAFF, 1987; 1995; CEFAS, 2000; 2003a; CEFAS, 2005; Feist *et al.*, 2008). Some locations, such as Liverpool Bay and Dogger Bank, that have heavy and complex loads of natural and anthropogenic pollutants (MAFF, 1990; Laslett, 1995; Camacho-Ibar & McEvoy, 1996; Langston *et al.*, 1999) show biomarker profiles characterised by high EROD activity (CEFAS, 2003c; Kirby, 2003), high frequency of DNA adducts (Lyons *et al.*, 2000; CEFAS, 2003b), and the prevalence of carcinogenic processes (Feist *et al.*, 2004; Feist & Stentiford, 2005;

Stentiford *et al.*, 2009). Furthermore the trends reported in the incidence of these biomarkers over time are congruent with changes in environmental quality (Jones, 2006) and somatic concentration of pollutants (Franklin, 2005).

Of all diseases encountered in dab, liver nodules and other hepatic tumorigenic processes have attracted most attention. The causality between many pollutants and liver lesions has been extensively studied (Mix, 1986; Stein *et al.*, 1990; Johnson, 1998; Järup, 2003; Waisberg *et al.*, 2003; Koehler, 2004), thereby rendering liver lesions important and informative indicators of pollutant exposure for which a set of guidelines has been standardised (Feist *et al.*, 2004) together with quality assurance procedures (BEQUALM, 2009).

Nevertheless, the elevated incidence of tumours in some locations considered pristine, like Cardigan Bay sites, remains unexplained (Lyons *et al.*, 2006). The use of a mobile species entail problems in linking biomarker responses to collection location (CEFAS, 2003a; van der Oost *et al.*, 2003), as fish heavily exposed to pollution in one area could migrate to a non-polluted area and, upon collection for biomonitoring assessment, create a false impression that toxicants operate in that location. Conversely, if healthy fish from non-polluted areas have recently migrated into a target contaminated site, their lack of long term responses will dilute the signal of pollution exposure, and the biomonitoring assessment will fail to report the true impact of local sources of pollution.

One effective way to identify marine population boundaries and connectivity, and the provenance of individual fish is to use genetic markers (Carvalho & Hauser, 1994; 1998; Hauser & Carvalho, 2008), especially markers exhibiting high levels of polymorphism and potential for individual assignment, such as microsatellites (Manel *et al.*, 2002). Microsatellites have also attracted attention from the ecotoxicology community (Brown *et al.*, 2001; Dimsoski & Toth, 2001) and have been employed in a few environmental pollution impact assessment studies (Berckmoes *et al.*, 2005; Maes *et al.*, 2005; Fratini *et al.*, 2008; Nowak *et al.*, 2009). In relation to such application, a suite of microsatellites for dab was developed (Tysklind *et al.*, 2009b). Weak but highly significant and temporally stable structure was found between North Sea and Irish Sea stations (Chapter 7). In this chapter, the utility of employing microsatellites in studies of biomonitoring in dab are examined in a detailed case by case analysis.

Firstly, the distribution of genetic diversity and changes in biomarker incidence will be evaluated to extract relevant biological relationships among samples. An assignment exercise

(Paetkau *et al.*, 1995; 2004; Rannala & Mountain, 1997) based on dab's population genetic data (Chapter 7) will then be performed to examine the putative origin of fish afflicted with liver nodules. The underlying rationale was based on the notion that if diseased fish at a particular location belonged to the same gene pool as healthy fish found at that location, the former would most likely be assigned back to the original sample. Conversely, if diseased fish were assigned elsewhere, then some process (i.e. migration or convergent selection) might be driving the genetic differences between healthy and diseased fish.

Pollution has been hypothesised to exert strong selective pressures on organisms which could lead to bottlenecks and reduction of genetic diversity (Bickham *et al.*, 2000; Belfiore & Anderson, 2001). Indeed, pollutants have been shown to impact allele richness and heterozygosity, both experimentally and in wild populations. Gardeström *et al.*(2006) evaluated the effects of flame retardants on exposed and control replicate populations of the copepod *Nitocra psammophila*, and detected significantly reduced heterozygosity in the exposed populations. Likewise, Nowak *et al.* (2009) reported a reduction of allele diversity and heterozygosity in midges, *Chironomus riparius,* after experimental exposure to tributyltin (TBT). In wild populations of flounder, *Platichthys flesus,* from different estuaries, Marchand *et al.* (2003) reported reduced heterozygosity in samples collected from polluted estuaries compared to the reference estuary, although the differences were small and not significant. Nonetheless, some field studies have sometimes reported the opposite. Theodorakis & Shugart (1997) found significantly higher levels of heterozygosity and diversity in mosquito fish, *Gambusia affinis*, living in a noxious cocktail of radionucleotides, heavy metals, PCBs and PAHs, than in those fish living in pristine environments. Similarly, Peles *et al.* (2003) also reported significantly higher levels of heterozygosity in earthworms, *Lumbricus rubellus,* living in heavy metal contaminated plots than in nearby control soils. The issue becomes even more complex when the individual heterozygosity relationship with pollution tolerance is studied. Larno *et al.*(2001) reported lower biomarker induction (EROD) in more heterozygous individual chub, *Leuciscus cephalus*; and Maes *et al.*(2005) reported reduced heavy metal load in more heterozygous eels, *Anguilla anguilla*. On the other hand, Bourret *et al.* (2008) found the opposite trend in yellow perch, *Perca flavescens,* with more heterozygous individuals carrying higher levels of heavy metals. These apparently contradictory patterns can be still be reconciled. For different species the detoxification and tolerance processes are different (Reichert *et al.*, 1998), in some cases the advantage may reside in being able to excrete the pollutant (Larno *et al.*, 2001; Maes *et al.*, 2005), while in others highly heterozygotic individuals may be able to tolerate higher levels of pollutants (Bourret *et al.*, 2008). In all

these cases, highly heterozygote individuals have a survival advantage which would produce more heterozygotic populations such as those found by Theodorakis & Shugart (1997) and Peles *et al.* (2003). The opposite trends observed in the experimental exercises (Gardeström *et al.*, 2006; Nowak *et al.*, 2009) could be explained by the selection exerted by a single toxicant selecting just for one genotype combined with the effects of reduction in population size in experimental trials.

Multilocus heterozygosity has often been hypothesised as an indicator of general individual fitness (Coulson *et al.*, 1998; Coltman *et al.*, 1999; Rowe & Beebee, 2005; Mainguy *et al.*, 2009) as inbred individuals are more likely to be homozygous for individual loci, both neutral and coding. The negative aspects of inbreeding emerge as an increasing number of recessive deleterious alleles become expressed in the homozygous state (Beebee & Rowe, 2004). Relationships between heterozygosity (either positive or negative) and general individual fitness in the form of survival, growth rate, parasite load, overall symmetry, or stress resistance have been studied widely(Danzmann *et al.*, 1988; Beaumont, 1991; Blanco *et al.*, 1998; Shikano & Taniguchi, 2002; Borell *et al.*, 2004; Blanchet *et al.*, 2009), although these relationships may vary between populations (Pogson & Fevolden, 1998). As reviewed before, several studies have also found heterozygosity to be relevant in pollution tolerance, thus, dab varying in average heterozygosity may perform differently after exposure to pollutants. Accordingly, the correlation between heterozygosity or inbreeding coefficient and various disease and parasite records will be examined here.

The high incidence of liver lesions across the whole tumorigenic process (Feist *et al.*, 2004; Stentiford *et al.*, 2009), together with the possibility of screening large numbers of individuals open the possibility of using dab as epidemiological models of cancer incidence in humans (Rotchell *et al.*, 2009). Here, the information available on the incidence of human cancer will be exploited to formulate hypothesis to be tested on the data set. Certain polymorphic genes may play a crucial role on whether the tumorigenic process progresses onto detectable pathologies (Nebert *et al.*, 1996; Nock *et al.*, 2006; Ryk *et al.*, 2006; Wu *et al.*, 2006; Fasching *et al.*, 2009; Whibley *et al.*, 2009), thus certain cancers are more prevalent in some human populations than others(Bhisey *et al.*, 2003; Chikako *et al.*, 2005; Distelman-Menachem *et al.*, 2009), and differences in cancer development risk are found at the family level (Lynch & Smyrk, 1996; Pharoah *et al.*, 1997; Lynch *et al.*, 2009; Permuth-Wey & Egan, 2009). If the propensity of developing cancers is family-specific also in dab, then fish recorded with liver nodules might be more related to each other than the rest of non-affected individuals. The

familial relationships between those fish exhibiting liver nodules will also be examined and compared to the average relatedness of all fish captured in the same area.

8.3 Materials and Methods:

Samples for the genetic analysis of individual dab were collected following the CSEMP assessment of biomarkers on four consecutive years (2005-2008) in several stations covering three basins: the North Sea, the English Channel, and the Irish Sea.

8.3.1 Biomarker data:

The results of the biomarker screening (i.e. individual disease records) were provided by CEFAS. The samples considered for genetic analysis here were a mixture of both "comprehensive" and "externals only" screened fish (Table 8.1) as defined in the introduction. Three samples (NeD07, Rye07 and Liv07) were analysed at the "comprehensive level" (i.e. all biomarkers were included). For these three samples age data read from otoliths was also provided by CEFAS. The rest of the samples were evaluated only at "externals level". Two further levels of analysis were evaluated for all samples: at the sample level, and at the individual level.

Disease profiles were handled as in Stentiford *et al.* (2009):

1) "externals": five grossly visible diseases recorded for all fish: lymphocystis, epidermal papilloma, skin ulceration, skin hyperpigmentation and liver nodules (Table 8.1).
2) "Comprehensive": composed of the "externals" plus an extra five categories described in Stentiford *et al.* (2009) summarising 32 different liver lesions classified according to their severity from Category 1 (non-specific) to Category 5 (malignant neoplasms) (Figure 8.2).

Individual disease records of all genotyped individuals were first transformed into presence or absence (1 or 0) of all aforementioned categories and within-sample prevalence was calculated as average among individuals.

Table 8.1: Sample information

Location and abbreviation (**Abv.**) used in the text. **Basin** corresponds to the general division found with genetic methods among dab samples. **Region** corresponds to the allocation of samples according to disease incidence (Stentiford *et al.*, 2009). **Lat. Long.** = coordinates where samples were collected. **Sample size** = number of individual fish genotyped for each location and year, numbers in *italics* are the number of individuals with no missing data. Disease coverage indicates whether the genotyped fish were assessed only for external and grossly visible diseases (**Ext.**) or where fully analysed for internal histopathology (**Comp.**)

						Sample Size								Disease coverage			
Location	**Abv.**	**Basin**	**Region**	**Lat.**	**Long.**	**2005**		**2006**		**2007**		**2008**		**2005**	**2006**	**2007**	**2008**
North East Dogger	NeD	NS	NSO	55.3000	2.8970	81	*64*	92	*87*	183	*180*	92	*72*	Ext.	Ext.	Comp	Ext.
Amble	Amb	NS	NSI	55.2669	-1.2543	42	*34*	86	*77*	47	*44*	95	*85*	Comp	Comp	Ext.	Ext.
Off Flamborough	Off	NS	NSI	54.2453	0.4985	21	*17*	37	*19*	48	*47*	94	*91*		Comp	Ext.	Ext.
Indefatigable Bank	Inf	NS	NSI	53.5567	2.0820	-		59	*46*	48	*48*	100	*96*	-	Ext.	Ext.	Ext.
Rye Bay	Rye	NS	EC	50.7790	0.7305	-		92	*91*	178	*174*	96	*80*	-	Ext.	Comp	Ext.
Lyme Bay	Lyb	NS	EC	50.6143	-2.9303	46	*41*	77	*69*	50	*50*	50	*42*	Ext.	N/A	Ext.	Comp
South Cardigan Bay	Scb	IS	ISS	52.1816	-4.4978	-		99	*98*	-		-		-	Ext.	-	-
Inner Cardigan Bay	Inc	IS	ISS	52.3000	-4.2725	99	*79*	39	*36*	49	*47*	30	*30*	Ext.	Comp	Ext.	Comp
Red Wharf Bay	Rwb	IS	ISN	53.7446	-4.1828	-		50	*50*	50	*50*	-		-	Comp	Ext.	-
Liverpool Bay	Liv	IS	ISN	53.4720	-3.6985	-		47	*42*	183	*173*	95	*41*	-	Ext.	Comp	Ext.
St. Bees Point	Stb	IS	ISN	54.5118	-3.7938	30	*18*	95	*92*	48	*48*	96	*77*	Ext.	Ext.	Ext.	Ext.
Overall		-				319	*253*	773	*707*	884	*861*	748	*614*				

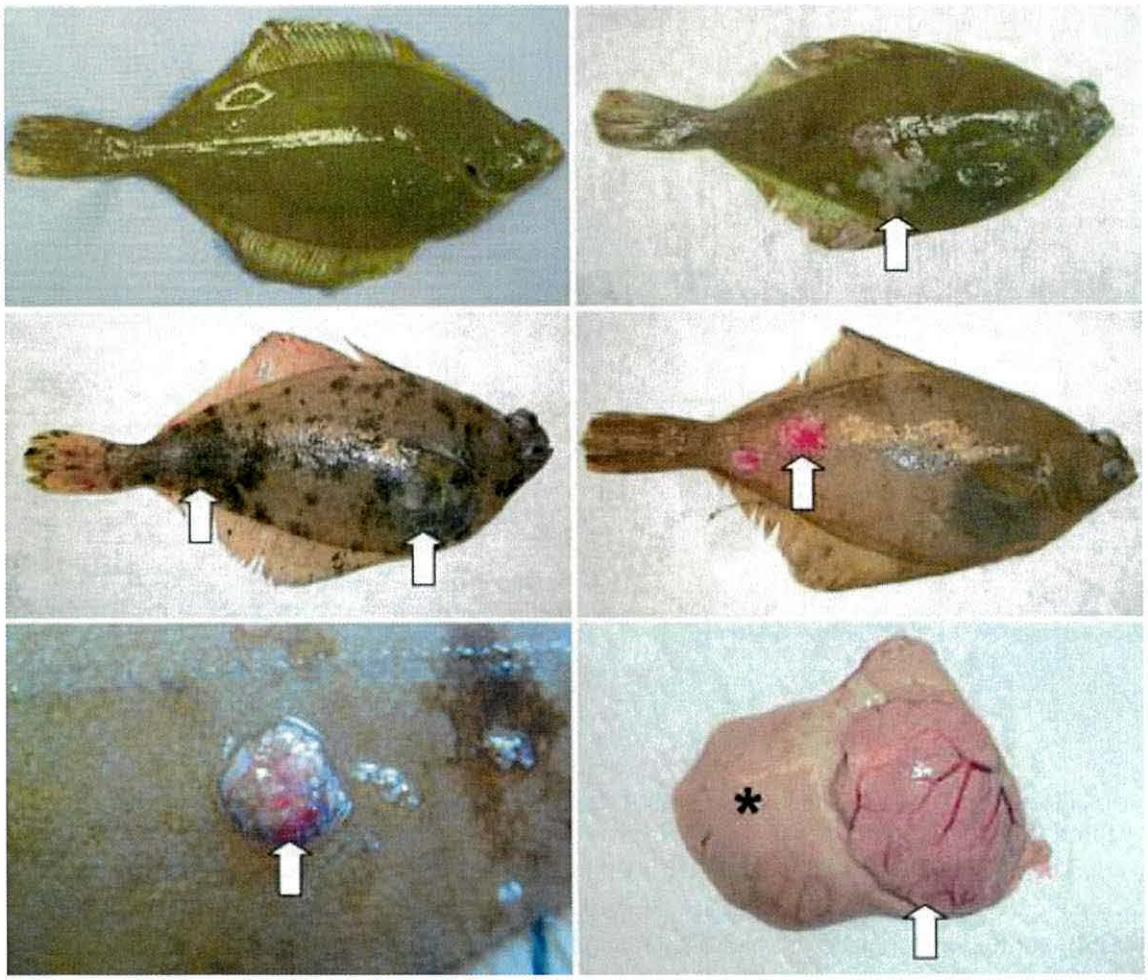


Figure 8.1: Grossly visible diseases of dab, *Limanda limanda*.

Reproduced from Stentiford et al, 2009. Top left = healthy fish with no visible diseases; Top right = Epidermal papilloma (EP); Centre left = Skin hyperpigmentation (HYP); Centre right = Skin ulceration (U); Bottom left = Lymphocystis (LY); Bottom right = Liver nodule (LN) and remaining apparently normal liver (*).

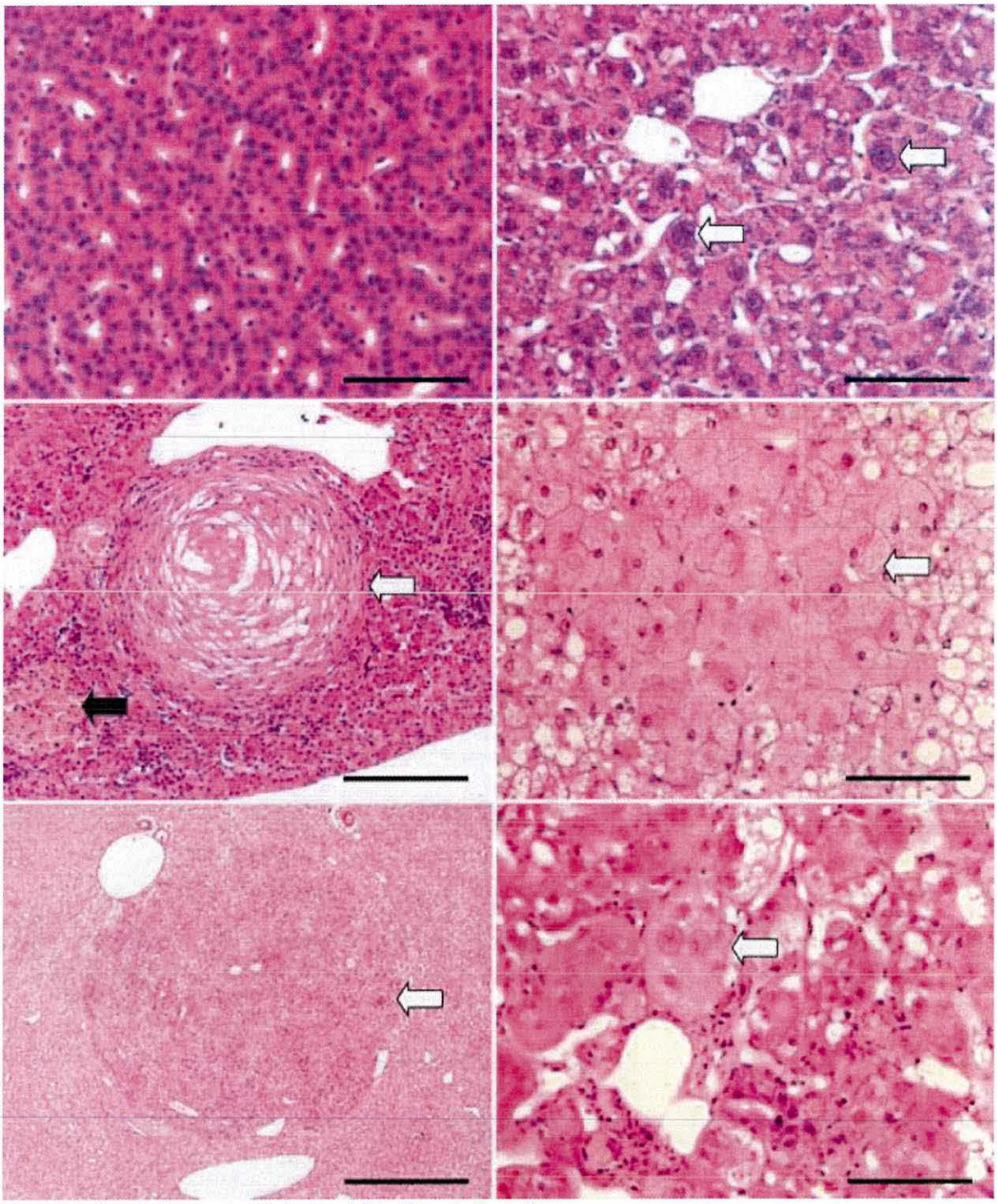


Figure 8.2: Liver lesions and pathologies of dab, *Limanda limanda*.
Reproduced from Stentiford et al, 2009. Top left = Normal liver with no abnormality detected; Top right = Nuclear pleomorphism (Cat1); Centre left = Granuloma (white arrow) and melanomacrophage (black arrow) (Cat2); Centre right = Focus of cellular alteration (Cat3); Bottom left = Hepatocellular adenoma (Cat4); Bottom right = malignant neoplasm (Cat4). Scale bars: top left and right (100 µm); middle left and bottom left (200 µm); middle right and bottom right (50 µm).

8.3.2 Genetic profiling:

Individual genotype data was composed of 16 microsatellite loci (Tysklind *et al.*, 2009b) (see Chapters 6 and 7 for details). To understand the distribution of genetic diversity of dab around the British Isles and the connectivity between samples, the results of several genetic analysis were re-examined, namely, neighbour-joining genetic distance trees (Felsenstein, 1989), GENELAND plots (Guillot *et al.*, 2005; 2005; 2008; Guillot, 2008), several forms of multivariate analysis (sample correspondence analysis and spatial principal component analysis) (Jombart, 2008; Jombart *et al.*, 2008; 2009), and admixture proportions at population level (Bertorelle & Excoffier, 1998; Dupanloup & Bertorelle, 2001). The methodology of the analysis of genetic data is described in Chapter 7.

8.3.3 Sample level analysis:

8.3.3.1 Evaluation of disease incidence:

The multivariate statistical package PRIMER™6.1.6 (Clarke & Warwick, 2001) was employed to analyse the biomonitoring data. Factors, which are used in some of the analysis, can be introduced *a priori* into the analysis. Three different factors were considered here: "Basin" as defined in the population genetics chapter (North Sea vs. Irish Sea, where English Channel samples were considered North Sea), "Year" (2005-2008), and "Region" as defined by Stentiford *et al.* (2009): North Sea Offshore, North Sea Inshore, English Channel, Irish Sea South, Irish Sea North. The significance of each factor at explaining biomarker variability among samples was assessed with an analysis of similarity (ANOSIM) with 10,000 permutations.

A principal component analysis (PCA) was performed to evaluate the contribution of each disease to the distribution of biomarker data. A resemblance matrix based on Euclidean distance was calculated and then used to produce non-metric multi-dimensional scaling (MDS) plots onto which resemblance boundaries were overlaid. The incidence of each of the diseases was incorporated into the MDS in the form of bubble plots where larger bubbles represent a higher frequency of the disease. The resemblance matrix was also used to generate dendrograms of sample similarity, where node robustness at the 50% level was assessed by the SIMPROF test with 10,000 simulations of 50,000 permutations each.

8.3.3.2 Correspondence between distribution of disease profiles and genetic diversity:

In order to assess the relationship between disease prevalence and genetic structuring, the Euclidean distance of disease prevalence between samples collected each year was compared

to genetic differentiation $\Theta'_{WC}/(1-\Theta'_{WC})$ (Rousset, 1997) of the same samples and tested with Mantel tests in IBD (Bohonak, 2002). Three levels were considered: first the whole data set (All) to assess general trends; then, single years (2005-2008) to check for changes among years; and finally, all samples from all years within the same basin (North Sea and Irish Sea) to assess if within-basin spatial-temporal changes in distances were similar. To avoid false significant correlations between genetic and disease distances, the English Channel samples were not included in either basin for the latter test. The significance of correlations was assessed with 100,000 randomisations of rows and columns of the Euclidean distance matrix.

Sharp changes in disease profiles and genetic composition were evaluated in BARRIER v2.2 (Manni *et al.*, 2004; Manni & Guérard, 2004). The programme generates cells around the sampling locations which have edges in common with other sampling locations, thus a network of connectivity between the geo-referenced locations is created. The coordinates of the sampling location were drawn and connection networks modified with empty cells (virtual points) to represent Great Britain eliminating the common edges between Irish Sea and North Sea, and between OfF and RyE, and LiV and LyB. Edges between StB and RwB and InC, and between Amb and NeD were purposely allowed. The two different distances, Θ'_{WC} and Euclidean distance of biomarkers, were plotted onto the map and compared against each other on a year by year basis.

8.3.3.3 Relatedness between diseased and non-diseased fish:

To test the hypothesis that the presence of grossly visible liver nodules are family specific the software STORM v1.1 (Frasier, 2008), which calculates relatedness (R) between individuals at several levels, was employed. Diseased groups were created with individuals ailing from liver nodules and average within-group relatedness (R_G) calculated. Only those locations with more than one diseased individual were considered for analysis. For the three samples with "Detailed" disease data, groups were also created for disease categories 3 to 5 (Cat3: Foci of cellular alteration; Cat4: Adenomas; Cat5: Carcinomas). In the latter, some individuals suffered from two or more categories and these were classified into whichever group was most severe (i.e. if an individual had both foci of cellular alteration and carcinomas, it was placed within Cat5).

The significance of the analysis was assessed by comparing the obtained results (R_G) with those produced by randomisations (R_G') in which a group of the same size as the diseased group was created by randomly sampling from all individuals from the location studied; therefore effectively evaluating the chance of obtaining the observed R_G values among a

subset of random individuals within the sample (R_G'). Only individuals fully genotyped were included in the non-diseased group, and genotypes were converted as suggested in the STORM manual to unique numbers from one to *n* from the smallest to the largest allele at each locus with the CONVERT excel sheet supplied with the software (available at http://www3.interscience.wiley.com/cgi-bin/fulltext/121494629/PDFSTART). The STORM manual stresses that the number of randomised iteration has to be smaller than the number of possible combinations of individuals (n(n-1)/2) within each group, thus for sample sizes around 50 individuals, 100 iterations were used (for example: a sample with 48 individuals of which three have liver nodules: number of combinations= [45(45-1)/2]+[3(3-1)/2]= 993 >>> 100), while 1,000 iterations were employed for those samples with nearly 100 individuals. The significance was assessed by counting the number of iterations (R_G') with a value above and below the obtained R_G.

8.3.4 Individual level analysis:

8.3.4.1 Relationships between Heterozygosity and bioindicator profiles:

Several estimators based on genotypic data have been devised to assess individual inbreeding level when pedigree information is not available: Coulson *et al.* (1998) devised a measure, d^2, which included theoretical molecular distance under a stepwise mutation model (Kimura & Ohta, 1978), though the relevance of the mutational model to the estimator has been questioned (Tsitrone *et al.*, 2001; Goudet & Keller, 2002). Amos *et al.* (2001) proposed another individual measure, internal relatedness (*IR*), based on the allele frequency-corrected "relatedness" of the two alleles at each locus averaged over loci. Positive *IR* values denote individuals whose parents are more related than average, while negative values are indicative of outbred ones. *IR* has been used extensively; nevertheless, Aparicio *et al.* (2006) pointed out problems with differences in loci allele diversity and rare alleles, and described yet another measure, homozygosity weighted by loci, *hL*, which takes locus allelic diversity into consideration.

Both *IR* and *hL* were calculated in STORM v1.1 (Frasier, 2008) for each individual in three large samples (over 100 fish) from 2007 (NeD07, Rye07 and Liv07), for which age and complete disease profiles were available. To avoid problems with loci exclusion (STORM manual), only individuals with complete genotypes were considered here. The relationships between both measures of inbreeding and disease (summarised into the 10 categories) and parasite records (both, individual parasite species, and sum of number of species infecting individuals) were analysed in MINITAB™. To assess whether more heterozygous individuals were in better

general condition, weight and length (standardised by age class), were used as a proxies of general fitness (Blanco *et al.*, 1998; Coltman *et al.*, 1998; David, 1998; Borell *et al.*, 2004) and compared to heterozygosity.

8.3.4.2 Assignment of diseased fish:

The assignment exercise was conducted in the software package GENECLASS2 (Piry *et al.*, 2004). First, fish genotypes were split into two groups, diseased (liver nodules) and healthy, for each location. Complete genotypes of healthy fish, which were the majority, at each location were used as reference samples. Genotypes of diseased individuals were then considered for assignment. The Bayesian multilocus assignment criterion (Rannala & Mountain, 1997) was used, and associated probabilities were calculated by simulating 10,000 individuals from the available genotypes (Paetkau *et al.*, 2004). The complete analysis was run independently for each year.

8.4 Results:

8.4.1 Population Level:

8.4.1.1 Evaluation of disease incidence:

There were strong differences in disease incidence between the samples (Table 8.2). The ANOSIM revealed that "Basin" was a weak descriptor for the distribution of biomarker data (Global R=0.094; p=0.058), as was "Year" (Global R=-0.044; p=0.810). "Region" on the other hand was the best factor fitting the biomonitoring data (Global R=0.431; p<0.001), thus "Region" was kept as factor for all subsequent analysis.

The PCA ordination plot of the grossly visible diseases (Figure 8.3) segregated sampling locations along the first principal component. The PCA1, PCA2, and PCA3 explained 81.2%, 9.0% and 6.2% of the variability of the data respectively. The Eigenvectors indicated that hyperpigmentation dominated the PCA1, which spaced North Sea samples from English Channel and Irish Sea; while skin ulceration was the main driver behind PCA2, which effectively separated coastal from offshore sites within the North Sea, and North and South sites within the Irish Sea.

The MDS plot with resemblance boundaries (Figure 8.3) and the Euclidean distance dendrogram (Figure 8.4) revealed significant clustering among samples within "Region" suggesting spatial and temporal stability of disease patterns for most locations. However, some temporal trends became apparent: there was an increase of hyperpigmentation

prevalence at NeD over time as depicted by the hyperpigmentation bubble plots (Figure 8.5). The StB07 sample stands out from the rest of Irish Sea (North) samples, which seems driven by low prevalence of skin ulceration and frequent liver nodules found in that location in 2007 (Figure 8.5). Of particular interest, are the Irish Sea (South) cluster composed of those samples collected in Cardigan Bay, all samples cluster together except for InC08 which instead seems more similar to English Channel or Irish Sea (North) samples. Also noticeable is Off06 which clusters with the Irish Sea South samples rather than the rest of North Sea (Inshore) samples.

Table 8.2: Sample "External" disease incidence.

The frequency per sample of five biomarker diseases. Sample name abbreviations are in Table 8.1. LY = Lymphocystis; U = Skin ulceration; EP = Epidermal papilloma; HYP = Hyperpigmentation; LN = Liver Nodules; The tables have been colour coded to facilitate reading: The stronger the orange and green the higher the value relative to other samples.

Sample	LY	U	EP	HYP	LN
NeD05	0.012	0.136	0.000	0.086	0.000
AmB05	0.048	0.024	0.000	0.167	0.000
LyB05	0.000	0.000	0.022	0.000	0.000
InC05	0.010	0.040	0.030	0.081	0.020
StB05	0.033	0.100	0.033	ournal	0.000
NeD06	0.000	0.141	0.000	0.087	0.000
AmB06	0.081	0.023	0.000	0.116	0.000
Off06	0.054	0.000	0.027	0.054	0.027
InF06	0.017	0.051	0.017	0.288	0.017
Rye06	0.011	0.011	0.000	0.022	0.000
SCB06	0.000	0.010	0.061	0.081	0.051
InC06	0.000	0.051	0.000	0.103	0.051
RwB06	0.040	0.040	0.000	0.020	0.020
Liv06	0.043	0.106	0.000	0.021	0.000
StB06	0.011	0.053	0.021	0.000	0.000
Ned07	0.000	0.104	0.000	0.306	0.071
Amb07	0.043	0.043	0.021	0.298	0.000
Off07	0.000	0.042	0.000	0.438	0.063
Inf07	0.000	0.000	0.083	0.333	0.021
Rye07	0.000	0.017	0.006	0.028	0.006
LyB07	0.000	0.000	0.000	0.000	0.000
InC07	0.000	0.041	0.020	0.061	0.061
RwB07	0.020	0.080	0.020	0.020	0.020
Liv07	0.000	0.082	0.016	0.027	0.016
StB07	0.021	0.021	0.042	0.000	0.188
NeD08	0.000	0.141	0.022	0.511	0.087
AmB08	0.095	0.011	0.011	0.168	0.000
Off08	0.021	0.000	0.021	0.181	0.000
InF08	0.000	0.010	0.030	0.190	0.020
Rye08	0.000	0.063	0.000	0.000	0.000
LyB08	0.000	0.000	0.000	0.000	0.000
InC08	0.000	0.033	0.033	0.000	0.000
Liv08	0.000	0.137	0.011	0.063	0.032
StB08	0.000	0.042	0.010	0.000	0.031

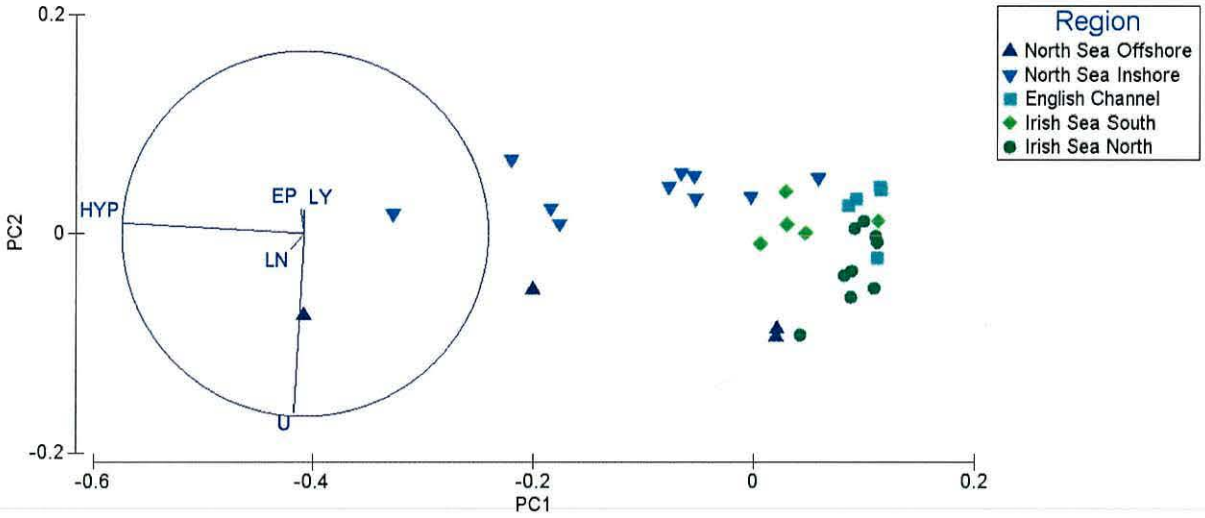


Figure 8.3: PCA of disease prevalence in dab, *Limanda limanda*.

Principal component analysis of five disease prevalence in dab samples collected from UK biomonitoring sites from 2005 to 2008 colour coded by **region**. The circle to the left represent vectors of each disease: HYP= hyperpigmentation; LY=Lymphocystis; EP= epidermal papilloma; U= skin ulceration; LN= Liver nodule.

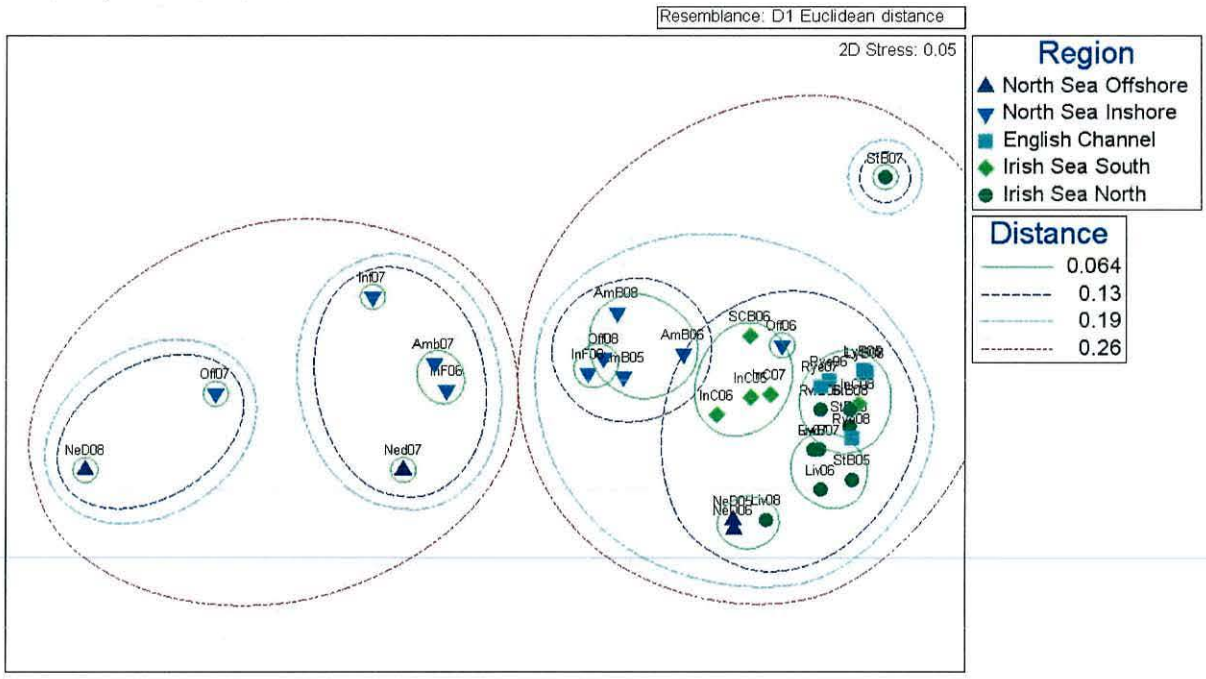


Figure 8.4: MDS plot of disease prevalence in dab, *Limanda limanda*.

Non-Metric Multi-Dimensional Scaling plot of disease prevalence data from dab collected in UK biomonitoring sites from 2005 to 2008 colour coded by region.

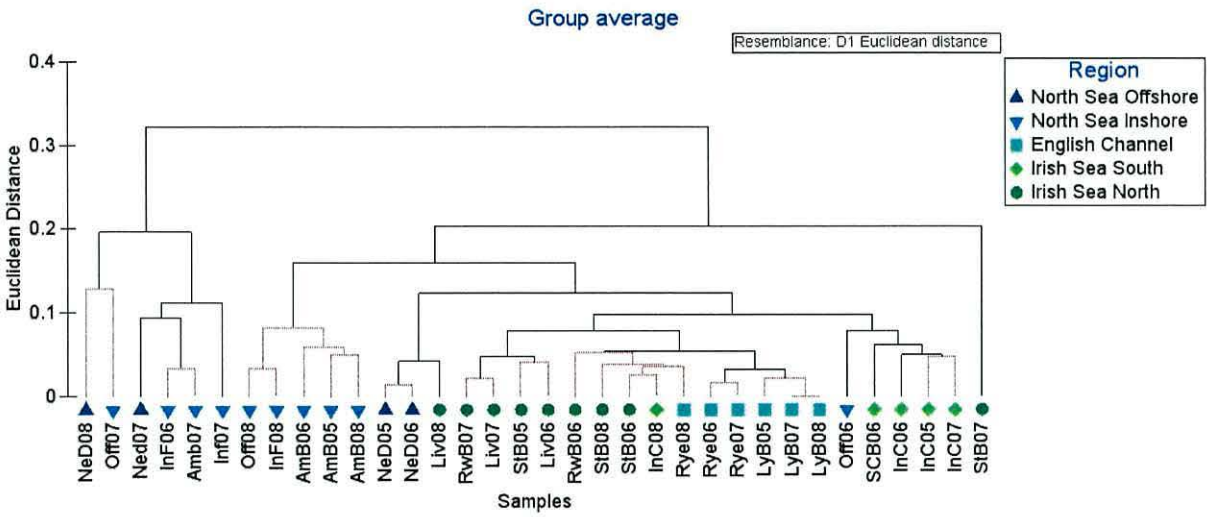
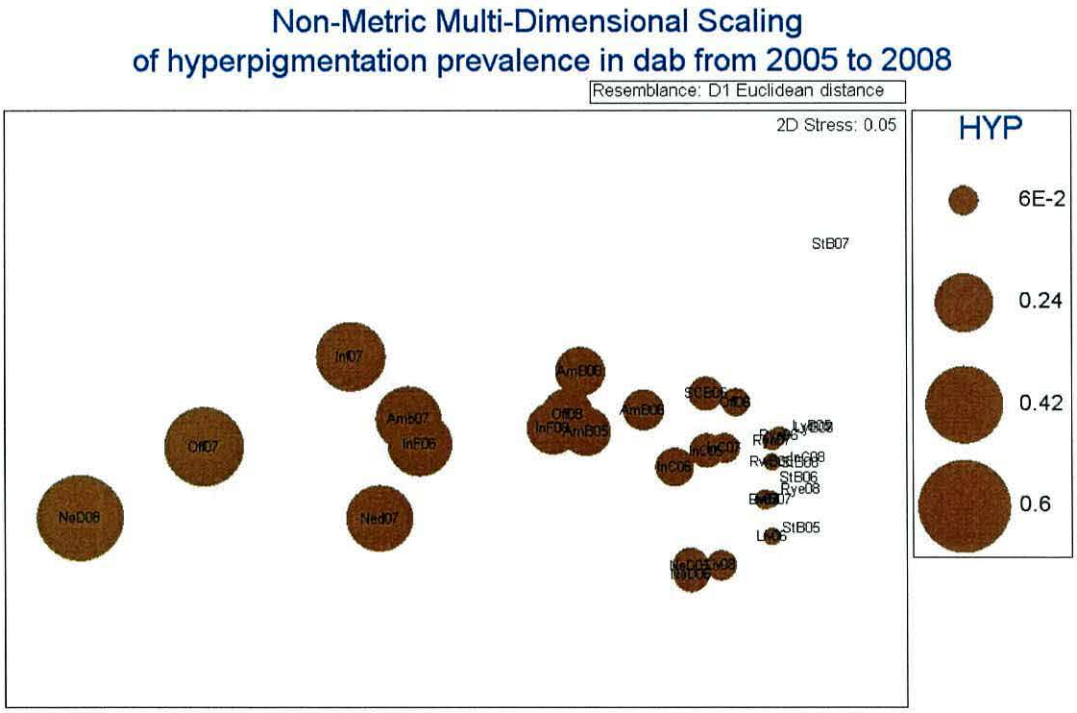


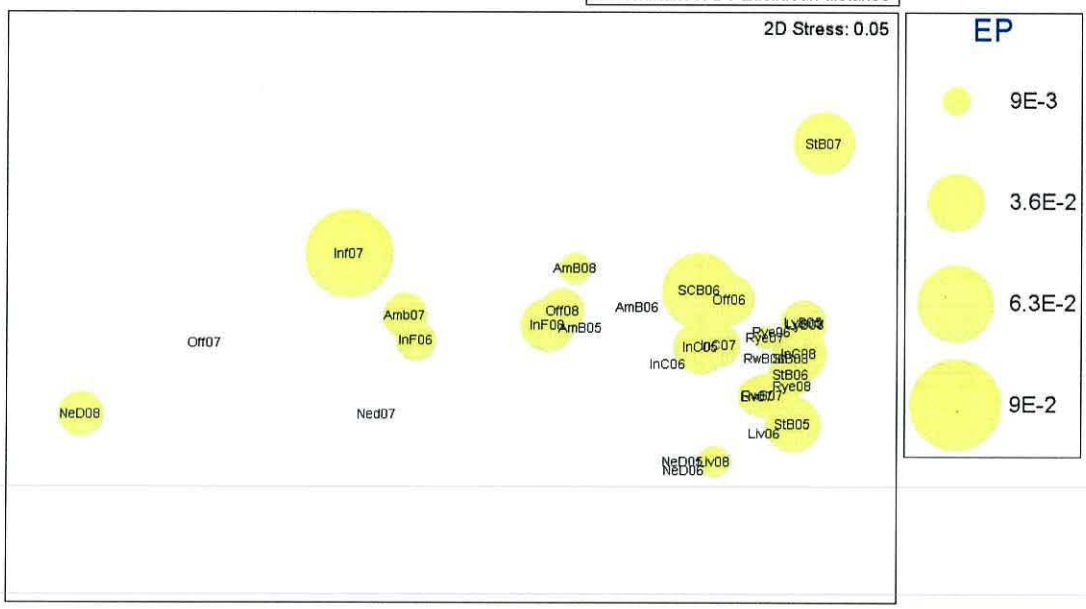
Figure 8.5: Dendrogram of disease profiles.

Dendrogram of samples based on Euclidean distance between disease profiles of dab samples collected in UK biomonitoring sites from 2005-2008. Nodes with less than 50% support after 50,000 permutations are depicted with dotted red lines. Samples are colour coded by region to simplify interpretation.



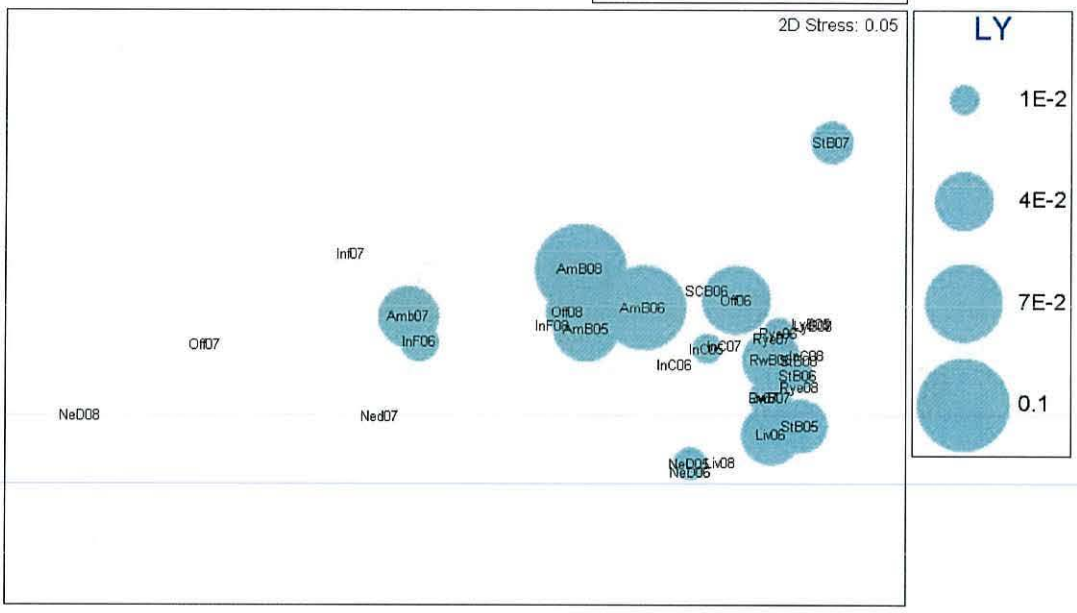
Non-Metric Multi-Dimensional Scaling of epidermal papiloma prevalence in dab from 2005 to 2008

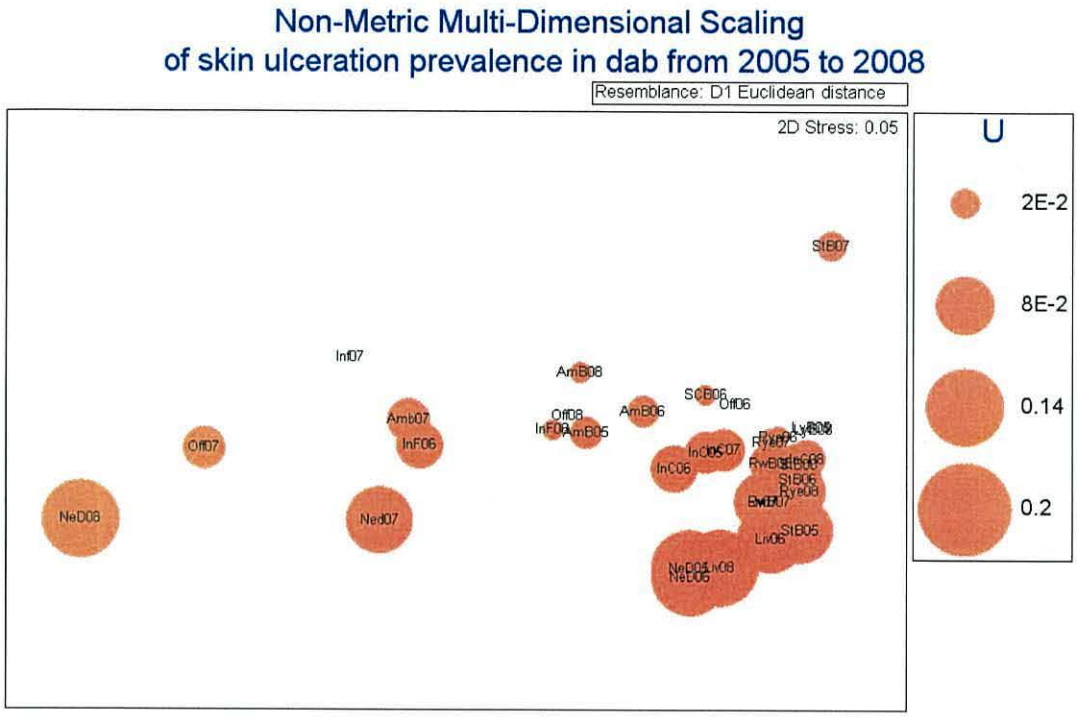
Resemblance: D1 Euclidean distance



Non-Metric Multi-Dimensional Scaling of lymphocystis prevalence in dab from 2005 to 2008

Resemblance: D1 Euclidean distance





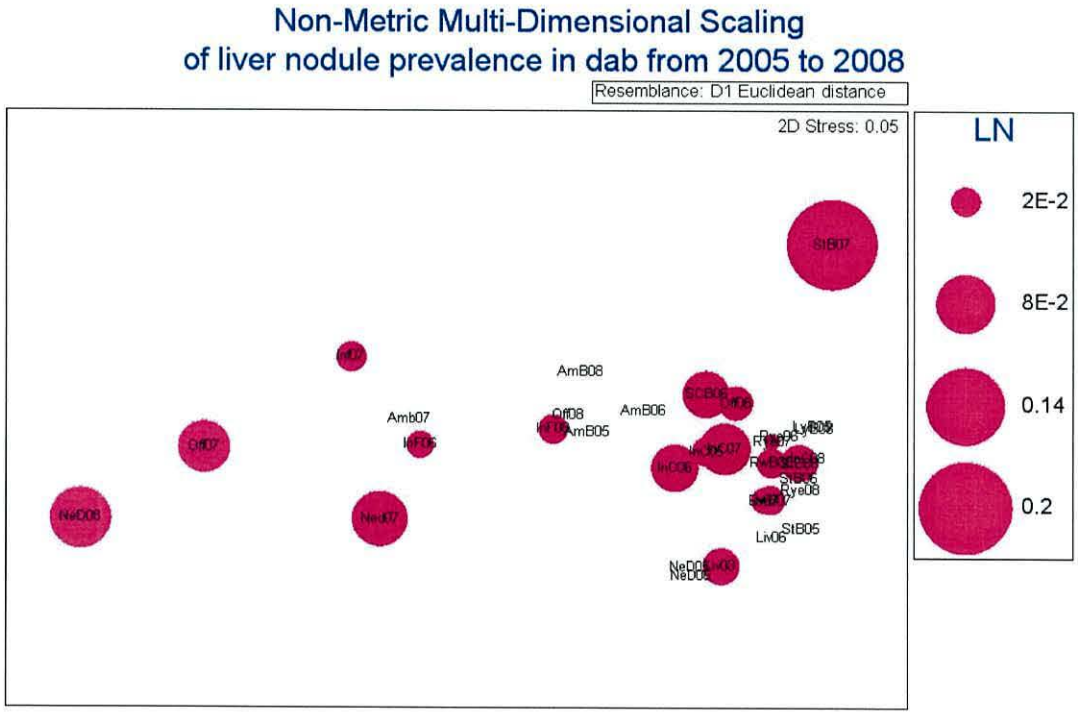


Figure 8.6: MDS bubble plots of single diseases

Non-metric multi-dimensional scaling of single diseases recorded in dab collected in UK biomonitoring sites from 2005-2008. The position of the samples is identical to that in Figure 8.4. The size of the bubble is relative to the frequency of the disease being analysed. A) Hyperpigmentation (HYP); B) Epidermal papiloma (EP); C) Lymphocystis (LY); D) Skin ulceration (U); E) Liver nodules (LN).

8.4.1.2 Correspondence between distribution of disease profiles and genetic diversity:

The Barrier analysis of the Euclidean distance highlighted the boundary between RyE and InF (The Calais-Dover strait) as the sharpest change in disease profile for 2005, 2006, and 2007. In 2008, the RyE-InF edge was ranked second after a boundary isolating NeD08 from the rest of the North Sea. The second ranked boundaries for 2005, 2006 and 2007 were Ned05-Amb05, InF06-rest of North Sea, and Stb07-rest of Irish Sea. Conversely, the strongest genetic changes were recorded at the other end of the English Channel: in 2006 and 2008 between InC and LyB, and in 2005 and 2007 between InC and the rest of the Irish Sea. The second boundaries for 2005 and 2008 separated InC from LyB05 and the rest of the Irish Sea respectively, while those of 2006 and 2007 separated NeD06 from coastal North Sea samples and isolated RwB07 from the rest of the Irish Sea (Figure 8.7).

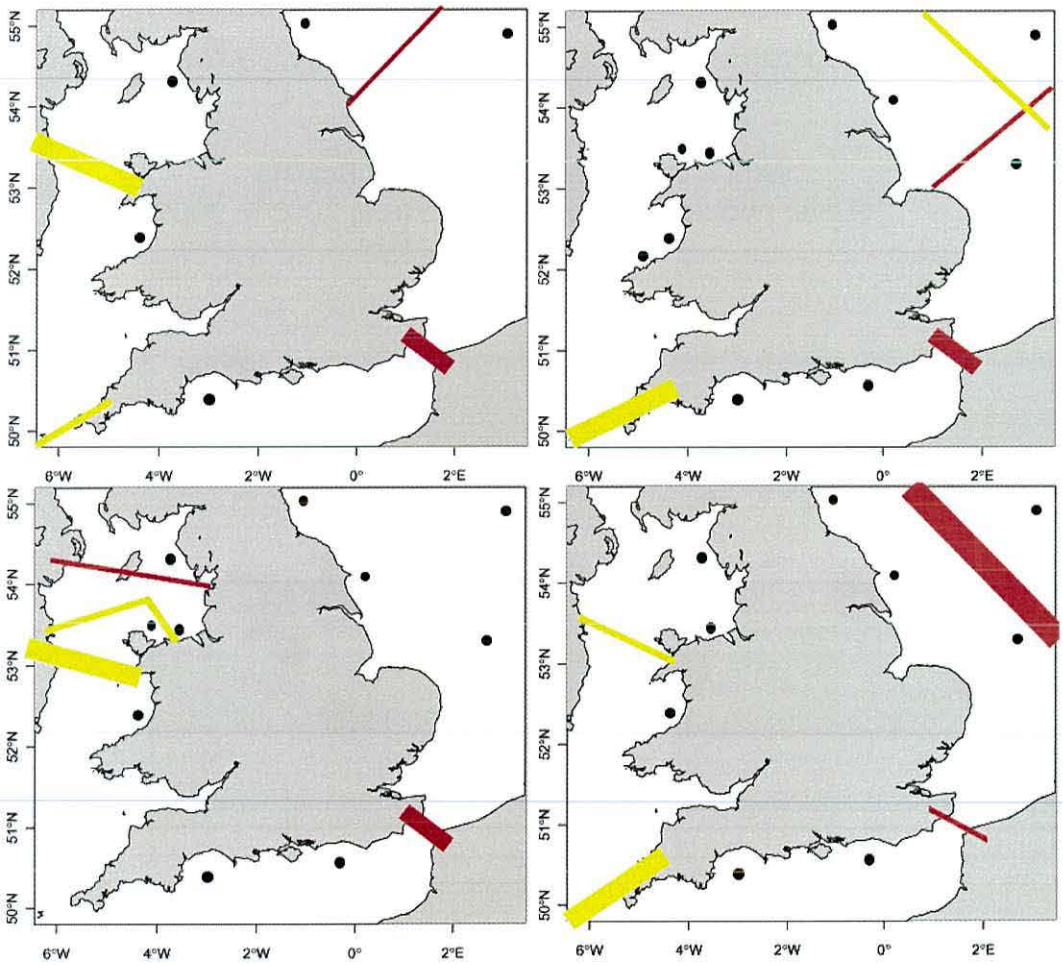


Figure 8.7: Strongest barriers to gene flow and disease prevalence.

First and second order ranking barriers for genetic distance (yellow) and disease Euclidean distance (red). First order barriers are thick lines, while second order ones are thinner. Sampling locations are represented by black dots. Maps are top left (2005); top right (2006); bottom left (2007); bottom right (2008). Note that barriers are placed in approximate locations. To enhance the accuracy of the location of a barrier an even more exhaustive sampling strategy would be needed.

The results of the Mantel tests (Table 8.3) suggest significant correlations between genetic (Θ'_{WC}) and Euclidean distances for 2006 and 2007. The relationship was not significant in 2005, 2008, for the whole data set, nor for either basin alone (North Sea or Irish Sea). The relationship between both distances was positively correlated, and nearly significant for North Sea (r=0.216; p=0.096), but not for the Irish Sea (r=-0.157; p=0.801). When the two distances were plotted against each other (Figure 8.8), two different relationships became apparent; first a positive relationship between both distances among North Sea and Irish Sea samples; and secondly, a cluster of dots with very low genetic differentiation but marked differences in disease profiles, which correspond to the North Sea-English Channel comparisons. As revealed by the BARRIER analysis, the English Channel samples were genetically more similar to the North Sea, but the disease incidence was rather low (Figure 8.3), and were very similar to the disease profiles found in the north of the Irish Sea. The discordant English Channel samples disrupted the relationship between both distances, and when the English Channel channel samples were removed from the Mantel analysis of all other samples, the relationship became highly significant (r=0.315; p=0.004).

Table 8.3: Correspondence between genetic distance and disease Euclidean distance.
Mantel test of correlation between genetic distance $\Theta'_{WC}/(1-\Theta'_{WC})$ and disease Euclidean distance. **r** = correlation value; **p** = associated probability after 100,000 randomisation of of the Euclidean distance matrix: p<0.05 denoted with *, p significant after Bonferroni correction (0.05/4 = 0.0125) are in **bold**.

	r	p	
2005	-0.370	0.808	
2006	0.505	**0.011**	*
2007	0.433	0.015	*
2008	0.085	0.293	

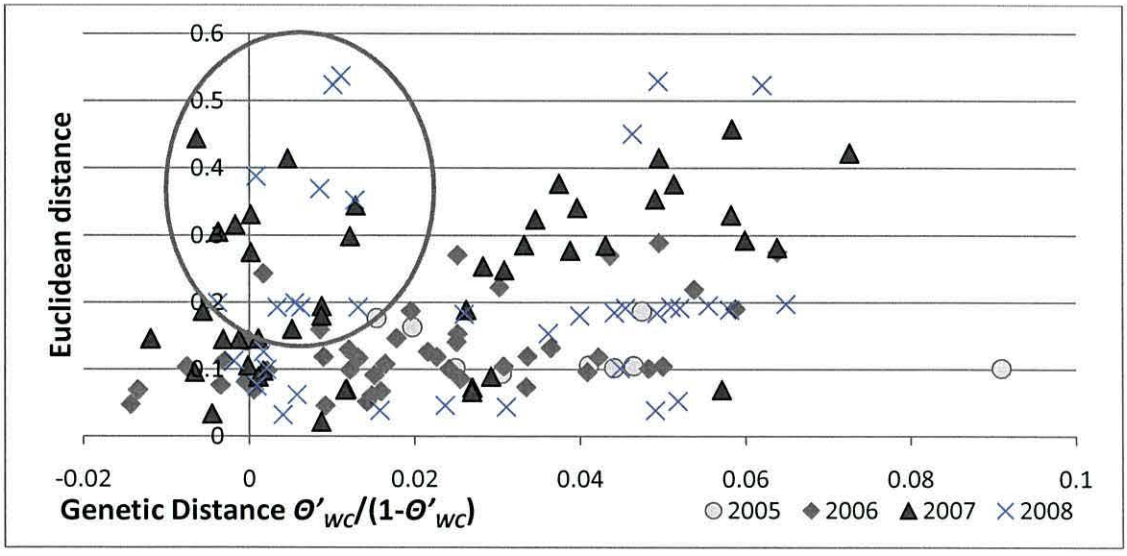


Figure 8.8: Genetic distance versus disease Euclidean distance.
Relationship between genetic distance $\Theta'_{WC}/(1-\Theta'_{WC})$ and Euclidean distance of disease prevalence in dab collected in UK biomonitoring stations from 2005-2008. The circle highlights those comparisons between North Sea and English Channel (low genetic distance but high disease Euclidean distance).

8.4.1.3 Relatedness between diseased and non-diseased individuals:

Only 10 samples fulfilled the condition of having more than one individual with liver nodules allowing the assessment of relatedness between diseased fish (Table 8.4). Most R_G values of both diseased and non-diseased groups were negative (average=-0.044) suggesting a general tendency towards non-relatedness, although most values again were not significantly different (within the 95% confidence interval of R_G' values calculated by randomisations) from those expected from a random grouping of individuals. The only significant value was that obtained from those individuals from InC07 with liver nodules group (R_G = -0.189; $p(R_G'<R_G)<0.01$), suggesting that individuals with liver nodules collected in InC07 were significantly less related to one another than any random combination of other fish collected in the same sample. Correspondingly, the inclusion of individuals with liver nodules in the randomised groups resulted in significantly lower R_G' than the observed R_G, (R_G=-0.011; $p(R_G'>R_G)$ =0.04).

Some R_G values were near to significance ($p<0.1$) at either end ($R_G'>R_G$ or $R_G'<R_G$): individuals with liver nodules from StB07 were the only group positively related and very few randomised values were higher than the observed value (R_G=0.024; $p(R_G'>R_G)$=0.070). In 2008 individuals with liver nodules from NeD and InF seemed less related among themselves than random draws of non-diseased fish (R_G=-0.061; $p(R_G'<R_G)$=0.066; R_G= -0.186; $p(R_G'<R_G)$=0.061).

Grouping individuals according to "Detailed" histopathology disease data (Cat3 to Cat5; Table 8.5) did not reveal any significant relationship patterns in NeD07 and Liv07. Nevertheless, non-diseased fish in Rye07 were significantly more related to each other (r= 0.013; $p(R_G'>R_G)<0.05$) than fish with foci of cellular alteration (r= -0.024; $p(R_G'<R_G)<0.1$).

Table 8.4: Average within-group relatedness (R_G) of dab, *Limanda limanda*, with liver nodules.

Samples in which more than one individual had grossly visible liver nodules (LN) were divided into two **groups**: reference and LN. LN groups are shaded in grey. The average pairwise relatedness (R_G) of group members is listed. The probability of obtaining values below (**%val<R_G**) and above (**%val>R_G**) the observed R_G, given the sample size and within-sample genetic diversity, are also provided. Significant values (p<0.05) are in denoted with a *.

	Group	R_G	%val < R_G	%val > R_G
2006	ScB06	-0.015	0.467	0.533
	ScB06LN	-0.039	0.307	0.693
	InC06	-0.037	0.520	0.480
	InC06LN	-0.076	0.410	0.590
2007	NeD07	-0.008	0.344	0.656
	NeD07LN	-0.020	0.309	0.691
	Off07	-0.023	0.080	0.920
	Off07LN	-0.017	0.570	0.430
	InC07	-0.014	0.960	0.040 *
	InC07LN	-0.189	0.000 *	1.000
	Liv07	-0.005	0.500	0.500
	Liv07LN	-0.043	0.328	0.672
	StB07	-0.033	0.150	0.850
	StB07LN	0.024	0.930	0.070
2008	NeD08	-0.011	0.257	0.743
	NeD08LN	-0.061	0.066	0.934
	InF08	-0.011	0.698	0.302
	InF08LN	-0.186	0.061	0.939
	StB08	-0.014	0.906	0.093
	StB08LN	-0.097	0.154	0.846

Table 8.5: Average within-group relatedness (R_G) of histopathology groups.

Three samples (NeD07, Liv07, and Rye07) were subdivided according to histopathology records: Cat3 = Foci of cellular alteration; Cat4 = Benign adenomas; Cat5 = Malignant neoplasms. The probability of obtaining values below (**%val< R_G**) and above (**%val> R_G**) the observed R_G, given the sample size and within-sample genetic diversity, are also listed. Significant values are denoted with a *.

Group	R_G	%val < R_G	%val > R_G
NeD07	-0.002	0.795	0.205
NeD07 Cat3	-0.014	0.295	0.705
NeD07 Cat4	-0.018	0.328	0.672
NeD07 Cat5	-0.052	0.262	0.738
Rye07	0.013	0.982	0.018
Rye07 Cat3	-0.024	0.087	0.913
Liv07	-0.006	0.445	0.445
Liv07 Cat3	0.007	0.682	0.318
Liv07 Cat4	-0.039	0.203	0.797

8.4.2 Individual Level:

8.4.2.1 Relationship between Heterozygosity and bioindicator profiles:

As reported by Mainguy *et al.* (2009), *IR* and *hL* were highly correlated to each other. Neither measure of inbreeding (*IR* and *hL*) was correlated with the calculated general fitness proxies (age-standardised weight and length) for any of the locations analysed (Appendix: Table B.1). Age had a significant relationship with *hL* in Rye07 ($r= -0.196$; $p<0.05$); Sex was also significantly correlated with both *IR* and *hL*; males were negatively correlated with inbreeding coefficients ($r\approx -0.220$; $p<0.05$), while the relationship in females was positive ($r\approx 0.220$; $p<0.05$). However, given the number of comparisons calculated and the lack of correlation at other sites, the effect of age and sex on *hL* is probably very weak or coincidental. In the NeD07 sample, Category 1 was significantly correlated with both measures of inbreeding ($r\approx 0.200$; $P<0.01$) (Appendix: Table B.1); however, Category 1 summarises non-specific liver pathologies, and when studied singly, no significant relationship was found with any individual pathology (data not shown).

Although not relevant from a heterozygosity point of view, many diseases were correlated with age at all sites (Appendix: Table B.2). Several diseases were also positively correlated with each other in NeD07 (e.g. skin ulceration, hyperpigmentation and liver nodules), as some individual fish often suffer from these pathologies simultaneously while others are relatively disease free.

8.4.2.2 Assignment of diseased fish:

The power (genetic differentiation assessed with 14 loci) of the assignment test was insufficient to exclude (with a $p<0.05$) alternative samples as potential sources of most individual fish. Nevertheless, the assignment test assigned many individuals with liver nodules back to the samples of origin, and for many others the second best option was also the sample of origin (Table 8.6), suggesting that these individuals are local fish and not migrants from other areas. When considering those individuals that do not assign back to the location of collection some interesting patterns emerge:

Table 8.6: Individual assignment of dab, *Limanda limanda*, with liver nodules

Results of the individual assignment of dab with liver nodules (LN) collected from UK biomonitoring sites from 2005-2008. The assignment was based on genotype data. Fish collected at each site with no visible liver nodules were used as reference. Each year was analysed independently (**2005-2008**). **Individuals** are listed on the first column and the probability of being assigned to each sample (Top row) is displayed below each sample name. Values in **bold** and highlighted in green reflect the best probability score of assignment to any of the samples considered. Values in light grey reflect probabilities below 0.05, indicating samples that can be excluded as a potential source population for the considered individual.

2005					
Individual	NeD05	Amb05	LyB05	InC05	StB05
InC05-109	0.116	0.003	0.170	**0.184**	0.010

2006										
Individual	NeD06	Amb06	Off06	InF06	Rye06	ScB06	InC06	Rwb06	Liv06	StB06
OfF06-994	**0.767**	0.404	0.308	0.500	0.683	0.687	0.298	0.520	0.762	0.681
InF06-D40	0.333	0.215	**0.618**	0.203	0.327	0.540	0.278	0.474	0.348	0.359
ScB06-001	0.090	0.003	0.004	0.027	0.039	**0.101**	0.002	0.044	0.003	0.034
ScB06-002	0.548	0.297	0.227	0.347	0.412	0.298	0.336	0.278	**0.555**	0.327
ScB06-014	0.821	0.792	0.405	0.692	0.866	**0.937**	0.732	0.874	0.910	0.876
ScB06-027	0.074	0.016	0.022	0.089	0.018	**0.199**	0.076	0.024	0.098	0.013
ScB06-029	0.870	0.713	0.651	0.382	0.769	0.880	0.667	0.923	0.585	**0.957**
InC06-113	**0.870**	0.659	0.348	0.240	0.524	0.811	0.650	0.635	0.534	0.554
InC06-137	0.142	0.021	0.020	0.252	0.105	**0.490**	0.280	0.177	0.368	0.362
RwB06-187	0.911	0.831	0.278	0.626	0.819	**0.969**	0.373	0.563	0.788	0.916

2008									
Individual	NeD08	Amb08	Off08	Inf08	Rye08	LyB08	InC08	Liv08	StB08
NeD08-011	0.082	0.282	**0.358**	0.337	0.025	0.095	0.028	0.009	0.023
NeD08-019	0.068	0.168	0.415	0.132	0.164	0.131	0.036	0.090	**0.555**
NeD08-026	0.977	0.938	0.899	0.864	0.811	**0.978**	0.219	0.758	0.928
NeD08-032	0.197	0.553	0.669	0.465	**0.710**	0.318	0.151	0.040	0.124
NeD08-033	0.119	0.310	**0.460**	0.423	0.221	0.230	0.358	0.056	0.173
NeD08-034	0.391	0.354	0.426	**0.569**	0.501	0.546	0.274	0.066	0.392
NeD08-035	0.014	0.236	**0.310**	0.177	0.091	0.074	0.037	0.052	0.127
NeD08-071	0.047	0.089	**0.290**	0.224	0.082	0.137	0.163	0.021	0.083
InF08-049	0.079	0.012	**0.300**	0.177	0.020	0.012	0.048	0.056	0.017
InF08-092	0.226	0.056	**0.234**	0.137	0.084	0.049	0.063	0.225	0.218
LiV08-018	**0.846**	0.537	0.482	0.801	0.767	0.613	0.583	0.691	0.569
LiV08-023	0.705	0.926	0.727	0.836	0.895	**0.932**	0.640	0.571	0.775
LiV08-032	0.506	0.719	0.665	**0.918**	0.650	0.710	0.319	0.600	0.665
StB08-015	0.213	0.353	0.403	0.276	0.113	**0.452**	0.101	0.212	0.226
StB08-028	0.211	0.104	0.077	0.147	0.046	0.170	**0.234**	0.073	0.187
StB08-056	0.679	0.407	0.787	0.179	0.734	0.605	**0.836**	0.439	0.622

Table 8.7 (cont.)

2007										
Individual	NeD07	Amb07	OfF07	InF07	Rye07	LyB07	InC07	Rwb07	Liv07	StB07
NeD07-1034	0.605	**0.659**	0.152	0.055	0.584	0.399	0.065	0.181	0.256	0.504
NeD07-1040	**0.371**	0.168	0.201	0.043	0.217	0.367	0.141	0.282	0.224	0.123
NeD07-1053	**0.278**	0.238	0.022	0.064	0.205	0.076	0.029	0.098	0.219	0.114
NeD07-1054	**0.994**	0.748	0.959	0.960	0.990	0.891	0.843	0.528	0.931	0.834
NeD07-1055	**0.964**	0.642	0.365	0.617	0.794	0.933	0.816	0.480	0.888	0.384
NeD07-1058	0.091	**0.441**	0.187	0.029	0.104	0.311	0.051	0.021	0.068	0.089
NeD07-1059	0.978	0.945	0.858	0.880	0.962	**0.985**	0.772	0.333	0.932	0.697
NeD07-1076	0.063	0.045	0.010	0.038	0.022	**0.245**	0.010	0.005	0.084	0.079
NeD07-1077	**0.255**	0.225	0.027	0.017	0.025	0.045	0.048	0.050	0.101	0.042
NeD07-1083	**0.990**	0.884	0.767	0.886	0.940	0.929	0.728	0.541	0.936	0.862
NeD07-1086	**0.137**	0.123	0.008	0.121	0.077	0.078	0.015	0.000	0.046	0.028
NeD07-1200	0.117	0.128	0.011	0.095	0.071	**0.143**	0.002	0.017	0.017	0.048
NeD07-1205	0.063	0.039	0.077	**0.499**	0.037	0.299	0.066	0.005	0.108	0.072
OfF07-1463	0.422	**0.637**	0.322	0.458	0.290	0.517	0.165	0.616	0.553	0.461
OfF07-1477	0.159	0.250	0.155	0.042	**0.392**	0.228	0.225	0.183	0.250	0.280
OfF07-1478	0.778	0.582	**0.869**	0.417	0.711	0.738	0.418	0.270	0.702	0.343
InF07-939	0.301	0.243	0.126	0.123	0.260	0.276	0.311	0.313	**0.570**	0.204
RyE07-1529	0.687	0.832	0.358	0.334	0.809	0.790	**0.856**	0.360	0.532	0.579
InC07-060	0.362	0.338	0.289	0.043	**0.386**	0.269	0.090	0.380	0.343	0.374
InC07-071	0.262	0.217	0.419	0.190	0.679	0.613	0.485	0.658	**0.735**	0.510
InC07-078	0.152	0.068	0.034	0.098	**0.222**	0.175	0.138	0.032	0.073	0.131
RwB07-111	0.449	0.187	0.100	0.176	0.283	0.055	0.164	0.082	0.400	**0.545**
LiV07-212	0.300	0.228	0.248	0.218	0.158	**0.351**	0.207	0.097	0.316	0.224
LiV07-321	0.202	0.015	0.085	0.152	0.077	0.153	0.119	0.078	**0.368**	0.007
LiV07-359	0.304	0.350	0.074	0.433	0.627	0.228	0.289	0.349	**0.680**	0.514
StB07-482	0.428	0.198	0.334	0.176	0.324	0.645	0.320	0.427	**0.719**	0.586
StB07-483	0.124	0.063	0.073	0.104	0.104	**0.372**	0.089	0.075	0.332	0.186
StB07-492	0.013	0.037	0.059	0.159	0.030	0.138	0.052	0.178	**0.224**	0.029
StB07-505	**0.897**	0.728	0.383	0.717	0.692	0.740	0.679	0.275	0.591	0.465
StB07-506	0.823	**0.876**	0.373	0.546	0.793	0.830	0.214	0.374	0.719	0.092
StB07-510	**0.253**	0.041	0.040	0.059	0.124	0.003	0.029	0.016	0.088	0.052
StB07-512	0.890	0.214	0.212	0.627	0.763	**0.962**	0.603	0.661	0.915	0.421
StB07-513	0.574	0.147	0.408	0.096	0.619	**0.750**	0.125	0.384	0.660	0.363
StB07-529	0.309	**0.434**	0.323	0.093	0.224	0.308	0.304	0.167	0.364	0.100

The only individual collected with liver nodules in 2005, was assigned back into InC. In 2006, all fish collected with liver nodules in the Irish Sea (ScB, InC, and RwB) were assigned to ScB06 (or it was the second best option with comparable scores to the first option). There was one exception (individual ScB06-002) which had a much higher probability of belonging to Liv06 than other Irish Sea locations. Interestingly, for one individual (ScB06-001) most other sites

could be excluded as potential sources, reinforcing the perception of genetic cohesion between diseased and non-diseased fish in ScB06.

In 2007, most fish collected in NeD with liver nodules were assigned back to NeD (or was a very close second best fit). Three fish had low probabilities ($p<0.1$) of being local fish, and were assigned to Amb, InF and LyB. The patterns in the Irish Sea were different from those in 2006. Two fish with liver nodules collected in InC were assigned to RyE (albeit with relatively low probability values), while a third individual was most likely a migrant from LiV. Similarly a fish collected in RwB was much more likely to be from StB or LiV than a local fish. Three fish were collected with liver nodules in Liv07: for two the first choice was LiV, while for a third fish LiV was a close second choice (after LyB). Most interestingly, the majority of the individuals reported with liver nodules in StB07 were assigned back into LiV (or was a very close second best). Three individuals had higher affinities elsewhere (NeD and AmB), but still the probability of belonging to LiV was higher than to StB for all three fish. StB07 was the only sample for which a positive (and near-significant) relatedness value was obtained.

Surprisingly, most of the fish with liver nodules from the North Sea in 2008 were assigned to Off08 (or were a close second best fit). In two instances, NeD could be rejected as a source for fish collected in NeD, and only one individual from NeD was likely to belong to the same pool as the rest of NeD fish. Similarly, none of the fish with liver nodules collected in LiV seemed local, but rather were most likely to be from LyB, Inf and NeD. Three fish suffered from liver nodules in StB08, but were more likely to be from InC and LyB.

8.5 Discussion:

8.5.1 Overview of the correspondence between distribution of disease profiles and genetic diversity:

The Barrier analysis revealed that the strongest boundaries between populations were different for genetic and disease data: the sharpest genetic boundary was apparent around the Southern Irish Sea (either North or South of InC), while the most noticeable change in disease profile was in the Dover-Calais strait. The difference between boundaries created a set of samples (RyE and LyB) that were genetically more similar to North Sea but exhibited disease profiles more in line of those from the Irish Sea. Once the English Channel samples were removed, the genetic and disease distances were highly correlated. The relationship is most probably not causative (i.e. the genetic composition of fish does not determine the pathological record) but coincidental (i.e. the same environmental and biological features that shape the genetic composition of dab in conjunction with local sources of contaminants produce the localised pattern observed in diseases profiles). The relationship between incidence of biomarkers, known history of pollution exposure, and genotype profile will be discussed on an area by area basis:

8.5.1.1 North Sea:

Dogger Bank samples are characterised by a high incidence of skin ulceration, liver nodules and frequent occurrence of all stages of the carcinogenic process (i.e. Cat3 to Cat5) (MAFF, 1995; CEFAS, 2003a; Feist & Stentiford, 2005; Feist *et al.*, 2008; Stentiford *et al.*, 2009). The presence of high levels of some heavy metals, particularly cadmium (Langston *et al.*, 1999), which has carcinogenic properties (Waisberg *et al.*, 2003), may be associated with the increased levels of disease incidence in Dogger Bank compared to other North Sea sites.

In the samples considered for genetic analysis in 2005 and 2006 there were no recorded liver nodules, though, this was because the individuals genotyped those years were from the smaller size classes (15-20cm) which are probably also the youngest and have not yet developed visible tumours(Baumann *et al.*, 1990). When the larger, presumably older size classes (non genotyped) were considered, the liver nodules frequency was 4.2% in 2005 and 7.3% in 2006, more in line with the values recorded for genotyped samples in 2007 (7.1%) and 2008 (8.7%).

The relatedness and assignment tests provide interesting insights into the source of individuals suffering from liver tumours across the North Sea. Most individuals with liver

nodules in 2007 were assigned back into NeD, and the relatedness test suggested that the groupings according to liver nodules (R_G) were random draws from the whole sample (R_G'). Together, both tests suggest that fish suffering from liver nodules are from the same gene pool as the rest of fish collected in NeD. Thus whatever is triggering the carcinogenic process should affect all fish at the Dogger Bank equally: either all fish have migrated together or the source of carcinogens is locally available. Given the temporal stability of the disease profile, the latter seems more probable. Three fish (out of 13) had low probabilities of belonging to the local population, and instead seem more related to coastal, or more likely given the low probability values, to some other unsampled population. Contrary to the apparent stability in 2007, the pattern changed completely in 2008 when most of the eight fish captured with liver nodules in NeD had much higher probabilities of belonging to Off08. For two fish NeD could even be rejected as a source population ($p<0.05$). Those fish with liver nodules collected in InF08 were also assigned back into OfF08. There were no obvious sample size differences between NeD08 and Off08 (72 & 91 after removal of incomplete genotypes) which could have biased the analysis. The two samples were slightly but significantly differentiated (Θ_{WC} = 0.001; Θ'_{WC}= 0.009; $p<0.05$) which could have help with the assignment test. Both groups of individuals with liver nodules from NeD08 and InF08 were near-significantly less related to each other and to the rest of the sample than expected by random draws from their respective samples (R_G = -0.061; $p(R_G'<R_G)$ = 0.066; R_G = -0.186; $p(R_G'<R_G)$ = 0.061), which also suggest that they do not belong to the same gene pool as the rest of the individuals collected at those locations. Two scenarios could explain such pattern: either a sudden increase of OfF fish migrating into NeD and exposed to carcinogens has occurred; or alternatively, there has been a change in the genetic composition of NeD non-liver nodules fish (i.e. the younger ones), while the older diseased fish are assigned to whichever is the closest population examined to the older NeD. The former scenario is unlikely as at least some local individuals would be expected among those with liver nodules (only one has a high second-best probability of belonging to NeD). In order to test the latter scenario effectively, age information would be needed: data that are not available here.

When analysing whole disease profiles, several samples did not match with those from other years collected at the same location. The OfF06 disease profile clustered with English Channel and Cardigan Bay samples; however, rather than clustering due to the presence of a particular disease, the uniting feature was the reduced prevalence of most diseases (Figure 8.5) (albeit there was one individual with liver nodules). Genetically, OfF06 was also an outlier; the sample correspondence analysis placed it outside the rest of North Sea samples

(Figure 7.8), the spatial principal component analysis suggested it was the most differentiated sample from the Irish Sea (Figure 7.9), and the GENELAND plots isolated OfF06 and InF06 from the rest of North Sea (Figure 7.5). The correspondence between both genetics and disease profiles strongly suggest that OfF06, or at least part of it, does not represent the same fish collective as other North Sea samples or samples obtained in Off Flamborough in other years. Instead, part of the sample may have recently emigrated into Off Flamborough from some other area with low disease prevalence that remained non-sampled for genetic analysis (as genetically OfF06 was different from all other samples). The OfF06 samples belonged to the "Comprehensive" group which are exhaustively studied for several projects, thus the recent migrant nature of OfF06 should be taken into account in those studies including that sample.

8.5.1.2 English Channel:

The English Channel experiences relatively low levels of pollutants (MAFF, 1990). Within the English Channel, RyE boasts the lowest disease incidence, and consequently, it is often considered as a reference site in biomonitoring programmes involving dab (MAFF, 1995; CEFAS, 2003a; Feist & Stentiford, 2005). In agreement with the published findings, the English Channel samples studied here were characterised by a generally low incidence of most diseases (Figure 8.5). The only fish with liver nodules genotyped from the English Channel over the course of four years (RyE07-1529) had a highest probability of belonging to Cardigan Bay, which normally experience a high incidence of liver nodules. Although the probability of being a local fish was also high, which would weaken the migrant hypothesis, the individual also suffered from skin ulceration and hyperpigmentation, which are rare in RyE, and thus supported the recent migrant hypothesis. The four individuals with foci of cellular alteration (i.e. beginnings of carcinogenic process) collected in Rye Bay in 2007 were significantly excluded as related to the rest of non-diseased fish, suggesting that they do not belong to the same gene pool as the rest of collected fish. When analysed in GENECLASS2, no individuals were assigned back to RyE, but instead to LiV, NeD and InF, and one had very low probability of belonging to any of the sampled populations (all $p<0.05$). Overall, the results suggest that there is a very low incidence of tumorigenic processes in dab from Rye Bay, and the few cases encountered are likely to be migrants from other areas. It is possible that the increased incidence of foci of cellular alteration reported in 2000 (CEFAS, 2003a) is due to increased migration from other areas.

Given the lack of genetically-local fish exhibiting liver nodules or foci of cellular alteration, it can be assumed that whatever is triggering the process elsewhere (i.e. Dogger Bank and

incidence of tumorigenic processes in dab from Rye Bay, and the few cases encountered are likely to be migrants from other areas. It is possible that the increased incidence of foci of cellular alteration reported in 2000 (CEFAS, 2003a) is due to increased migration from other areas.

Given the lack of genetically-local fish exhibiting liver nodules or foci of cellular alteration, it can be assumed that whatever is triggering the process elsewhere (i.e. Dogger Bank and Cardigan Bay) is absent in Rye Bay. Low exposure to carcinogens such as PCAs and PCBs may be proposed(Kirby *et al.*, 1999), rendering RyE an appropriate reference site for those sites with a North Sea genetic signal (North Sea, RyE and LyB in some years). However, Rye Bay was dominated by very young fish in 2007 (1 or 2 years old), and as reported by Baumann *et al.* (1990), age and most diseases (including liver nodules) are significantly correlated in all three sites. If fish in Rye Bay are always very young, it could explain the low disease incidence regardless of exposure to pollutants. Dab data from a previous sampling cruise (2004), in which all fish from all stations were aged, revealed that RyE was indeed the site with the youngest average age (Mean=2.93, S.E.=0.18), supporting the notion that Rye Bay is normally composed by young fish. The apparent lack of older individuals at Rye Bay raises the question of the fate of older individuals (>2-3 years): is local mortality and fish turnover very high? Or are they migrating elsewhere? The two hypotheses are discussed:

The low prevalence of diseases does not suggest mortality at young age; similarly mortality due to predation and fishing pressure are unlikely to be so markedly different among areas. The increased relatedness of non-diseased Rye Bay fish could suggest that the location may act as a nursery for dab: young cohorts of siblings occurring together for the first year or two and then dispersing or mixing with other cohorts born elsewhere. Given the age, disease and genetic relatedness, migration to other areas becomes the most plausible option. Rye Bay always clusters genetically with North Sea sites, so dab born in RyE could be migrating north as they mature and recruit into the North Sea sites. There were low but significant genetic differences between RyE and the northernmost locations in the North Sea for all years, thus from a genetic point of view, dab born in RyE are unlikely to recruit into Dogger Bank or northern coastal stations in large numbers. On the other hand, InF and RyE were genetically very similar in 2007 and 2008 so InF could be a receptive location for maturing RyE fish.

At the other end of the English Channel, the increase in apparent North Sea admixture in LyB over time reported in Chapter 7 (Figures 7.11 and 7.12 and Table 7.7) was not accompanied with an increase in disease incidence (Figure 8.3), suggesting that the increase in North Sea

genetic traits is more likely to have originated from within the English Channel (such as from RyE). Age data was not available for LyB, and so it was not possible to determine whether the low incidence of disease recorded at this site was related to young age or other processes.

To conclude, although Rye Bay genetically could be a reference site for sites in the North Sea, the young age of fish collected there confounds the interpretation of the low disease incidence detected, and thus compromises its utility as a realistic reference site.

8.5.1.3 Cardigan Bay:

In the late 1980's dab collected from Cardigan Bay were found to have the lowest disease prevalence of the Irish Sea (MAFF, 1987), and thus was considered as a relatively unpolluted reference site. Nevertheless, liver nodules started being reported from Cardigan Bay samples in the mid 1990's, although most of these liver nodules were not confirmed as hepatic adenomas, (MAFF, 1995) and there were no DNA adducts detected in outer Cardigan Bay in 1996 (Lyons *et al.*, 2000). Since then, Cardigan Bay dab have often been reported to suffer from a high incidence of liver nodules (CEFAS, 2003a) and confirmed adenomas and carcinomas (Feist & Stentiford, 2005; Lyons *et al.*, 2006; Feist *et al.*, 2008). A marked increase in DNA adducts from none being recorded in 1996 to levels comparable to Liverpool Bay in 2000 has been reported (CEFAS, 2003b). Concomitantly, in the current study, liver nodules were common in Cardigan Bay samples in 2005, 2006 and 2007. The assignment test suggested that, with few exceptions, most diseased individuals in 2005 and 2006 were assigned back to Cardigan Bay, and the relatedness test indicated that diseased fish were genetically a random draw of all fish collected in ScB06, altogether suggesting that the existence of tumours in dab is a local feature. The existence of tumours in dab captured both in summer and winter (Lyons *et al.*, 2006) is coherent with this view. On the other hand, the three individuals with liver nodules collected in 2007 were significantly less related to each other and other local dab than expected by random, and were assigned to other locations. Furthermore, two of them had skin hyperpigmentation which was very rare that year in InC (three in total) reinforcing the notion that they are biologically or geographically different to the rest of fish collected in the area that year. Although one of those individuals with liver nodules could be assigned to LiV with high probability, the other two had low probabilities overall for belonging to any sampled population, perhaps indicating that the population of origin was not genotyped.

InC08 was characterised particularly by the lack of any visible liver nodules. Although InC08 was clearly nested within the Irish Sea branch of genetic distance trees (Figure 7.2) and

correspondence analysis (Figure 7.8), GENELAND indicated that InC08 was different from other Irish Sea samples in 2008, and was placed together with LyB08 (Figure 7.7). Furthermore, the analysis of admixture (Table 7.7), which suggested no admixture for InC06 and InC07, indicated some influence from the North Sea in InC08. The influence is likely to proceed from the English Channel given the increase in North Sea genetic component, but reduced disease incidence typical of English Channel. Furthermore there were no recorded cases of hyperpigmentation in InC08, the lack of which is typical of English Channel sites, while fish from Cardigan Bay in other years had incidences of hyperpigmentation.

Much of Cardigan Bay has been protected as 'Special Areas of Conservation' (SAC) under the European Community Habitats Directive (http://www.jncc.gov.uk/page-1445) and is reported to have low levels of nitrogen compounds, phosphate, silicate, tributyl tin, and low concentration of polyaromatic hydrocarbons (PAHs) and organochlorinated pesticides (PCBs) (MAFF, 1990; CEFAS, 2000), some of which are known carcinogens (Hawkins *et al.*, 1988; Reichert *et al.*, 1998; Baumann, 1998; Srogi, 2007). In a four fish species assessment, PCBs were found to be an order of magnitude smaller in Cardigan Bay than in the Liverpool area (MAFF, 1990). Therefore the elevated occurrence of tumorigenic processes in dab is unexplained (Lyons *et al.*, 2006). The genetic data provided in the current study is not suggestive of consistent migration of diseased fish into Cardigan Bay; indeed, the converse appears to be the case as the low incidence of disease in 2008 is associated with immigrant genetic influence. Therefore a local source of carcinogens should be sought.

Wales has a large and increasing coverage of bracken fern, *Pteridium aquilinum*, (Pakeman *et al.*, 1996) a plant which has been demonstrated to have carcinogenic effect on many animals: 100% of rats fed on bracken developed multiple tumours (some of which were malignant adenocarcinomas) in the intestine and mammary glands(Evans & Mason, 1965); bladder neoplasias were also induced by bracken fern in guinea pigs, mice, and cattle (Bryan, 1977). The toxicity can be transferred through milk from mother (cows and mice) to offspring or consumers of milk products (Evans *et al.*, 1972; Alonso-Amelot *et al.*, 1998). Even the airborne spores have been found to be carcinogenic and the intensity of the DNA damage, in the form of DNA adducts, correlated with extent of exposure (Simán *et al.*, 2000). Associations between bracken exposure and the high incidence of gastric cancers in humans residing in North Wales have been reported (Galpin *et al.*, 1990), but others have questioned the strength of the causality of bracken in the development of human cancers (Wilson *et al.*, 1998). The toxic compound in bracken, ptaquiloside (Potter & Baird, 2000), is water soluble

and can be leached by rain from the fronds onto the topsoil and from there to rivers and water reservoirs, especially in heavy rainfall areas (Rasmussen *et al.*, 2003). Under neutral or slightly acid conditions, cold temperatures and low light exposure the compounds are quite stable (weeks and months) (Saito *et al.*, 1989) and thus, it is possible that heavy rainfalls after the sporing season (August-October) may carry large concentrations of ptaquiloside compounds into Cardigan Bay, creating DNA-adducts in dab and ultimately producing the tumours observed. Whether bracken leachate is responsible for the high incidence of dab tumours would need measurements of compound concentrations at sea after heavy rainfall and assessment of the toxicity of ptaquiloside on dab. Although bracken may be increasing in abundance in Wales (Pakeman *et al.*, 1996), it is still a native plant (Ashcroft & Sheffield, 1999), which makes the sudden increase in genotoxic damage shown in dab difficult to explain. Therefore other explanations will be explored.

Cardigan Bay is relatively free of industrial and domestic pollutants as the human population density is low and there are no large industrial centres around its shores. Nevertheless, large concentrations of lead, copper, zinc and cadmium have been reported in the northern tip and southern part of the Bay (Abdullah *et al.*, 1972). The existence of the heavy metals has been attributed to river runoff from mineral rich areas in Snowdonia, and mining activities in the XIXth century (Abdullah & Royle, 1972), which are then concentrated in the southern part of Cardigan Bay (underneath the InC and ScB sampling sites) by the local circulation pattern(Abdullah *et al.*, 1972). On the other hand, more recent work by Pearce & Mann (2006) studied the concentration of several heavy metals in the shells of razor shells (*Ensis siliqua*) and instead reported rather low levels of all contaminants in Cardigan Bay compared to Liverpool and Dulas Bays (North East shore of Anglesey). However razor shells live in the intertidal shore and the highest concentrations of heavy metals were reported in deeper waters (Abdullah *et al.*, 1972) and thus it is possible that dab are exposed to remnant high levels of heavy metals of natural origin or dating back to past mining activities. Heavy metals are known carcinogens (Johnson, 1998; Järup, 2003; Waisberg *et al.*, 2003) and could induce tumorigenic processes in local dab. Although exposure to heavy metals is a more plausible explanation to the high incidence of liver nodules in Cardigan Bay than fern leachate, these polluted sediments have been present since at least the 1970's when measured by Abdullah *et al.* (1972) which still does not explain the temporal pattern of carcinogenic processes detected in dab. The only way these settled pollutants could change the patterns of the diseases of the local population of dab suddenly is if sediments experienced atypical disturbance routinely from the 1990's. The start of aggregate dredging or energy reserves

exploitation could have such an impact by resuspending the sediments (Desprez, 2000; Hitchcock & Bell, 2004; Barrio Froján *et al.*, 2008), thereby leading to an increase in heavy metal exposure. However, neither dredging for aggregates nor oil reserves exploration is reported to occur in Cardigan Bay (http://www.cardiganbaysac.org.uk/?page_id=102)

Given the low incidence of cancer in 2008 together with the southern influx in genetic composition, perhaps the changes are due to population turnover of dab and differences in population tolerance. Hypothetically, the low incidence of diseases in the 1980's could be associated with a population adapted to the naturally high levels of heavy metals, which for whatever reason may have been replaced by a more sensitive population in the late 1990's. The reduced incidence in local fish in 2007 and 2008 may indicated another change in population tolerance.

8.5.1.4 Irish Sea North:

Dab have been studied in the Irish Sea for the presence of liver nodules since 1982 (MAFF, 1987), when Liverpool Bay was recognised as the most polluted area of the Irish Sea region. Ten years later, Liverpool Bay still recorded the highest levels of skin ulceration and liver nodules of all stations in England and Wales (MAFF, 1995). The probable cause of which were the high concentrations of mercury, cadmium, nickel, pesticides and other PCBs recorded in Liverpool Bay dab in the late 1980's (MAFF, 1990; Laslett, 1995), and sourced from the Mersey Estuary (Camacho-Ibar & McEvoy, 1996). Since then, considerable efforts have been made to improve the overall health of the estuary and an outstanding recovery of the Mersey has been reported (Jones, 2006). Mirroring the trend in improving ecosystem health in the Mersey, dab from Liverpool Bay have been showing a reduction in mean EROD activity (CEFAS, 2003c; Kirby, 2003), lower levels of mercury (Franklin, 2005), and were no longer the site with the highest prevalence in liver nodules in 2003 (Feist & Stentiford, 2005). In the latest published assessments, disease records of dab samples from LiV are comparable to those seen in other areas of the Irish Sea (Feist *et al.*, 2008; Stentiford *et al.*, 2009). Although the incidence of liver nodules in LiV was not outstanding in the samples genotyped in the current study, LiV represented a source of diseased fish collected in other areas, particularly in 2007. Fish collected in LiV in 2007 were likely to comprise local fish, though this was not the case in 2008.

RwB and StB monitoring stations are proposed to be rather cleaner that LiV was in the past, with lower incidence of diseases, albeit a few occasional instances of liver nodules (MAFF, 1987; MAFF, 1995; CEFAS, 2003a). The two individuals collected in RwB with liver nodules had

higher probabilities of originating from elsewhere, which suggests that although there are sources of contaminants nearby ((effluents from the nearby Parys Mountain can be detected in the shells of local molluscs, Pearce & Mann, 2006), generally the local population does not display liver nodules. Both cases were assigned to the prevailing source of liver nodules in the Irish Sea in each year (i.e. ScB in 2006, and StB or LiV in 2007). The published literature report that incidence of liver nodules in RwB is variable: no fish out of 272 showed liver nodules in 2003 (Feist & Stentiford, 2005), while there were 13 cases out of 419 in 2006 (Feist *et al.*, 2008), indicating potential differences in the proportion of migrants exhibiting liver nodules.

The problems with interpreting the role of migratory movements is the distribution of diseased fish (CEFAS, 2003b) are exemplified by those individuals displaying liver nodules in StB. No instances were recorded in 2005 or 2006 in the genotyped fish, and low incidence is recorded for those years (2.5%) (Feist *et al.*, 2008). In 2007, however, 18.8% of fish genotyped (48) were recorded with liver nodules. The extreme occurrence of liver nodules becomes even more interesting as most individuals were assigned to LiV and were close to being significantly more related to each other than random draws of the sample overall (R_G=0.024; p (R_G'>R_G) =0.070). Together, both tests strongly suggest that these fish have recently migrated from elsewhere, probably LiV. Again those individuals collected in 2008 were also assigned elsewhere, reinforcing the notion that the incidence of liver nodules in not something inherent to the local population in StB.

Overall the north of the Irish Sea seems like a highly dynamic area in term of dab movements; though some trends are apparent over four years: Liverpool and Cardigan Bays may be likely sources of diseased fish for other areas where the incidence of disease is lower (Red Wharf Bay and St.Bees Point), but such patterns will only be detected if genetic data, collected simultaneously with biomonitoring information, are available.

8.5.2 Heterozygosity and biomarkers:

There was no consistent evidence for any negative effects of homozygosity in any of the biomarkers considered. However, the results are not unexpected and are in agreement with other studies evaluating pollution impacts using microsatellites (Berckmoes *et al.*, 2005; Maes *et al.*, 2005). When analysing functional molecular markers (such as enzymes, coding single nucleotide polymorphism, and gene-linked microsatellites) heterozygosity may represent a real advantage if the polymorphism at the locus is functionally diverse. On the other hand, when analysing neutral markers (such as truly neutral allozymes and microsatellites) the relationship between microsatellite heterozygosity and individual performance

(heterozygosity at coding genes) is produced by genome wide linkage disequilibrium, such that high levels of homozygosity at the marker reflect equal levels of homozygosity at coding genes. However, this relationship is only expected to occur when inbreeding is evident (David, 1998; Tsitrone *et al.*, 2001). Although inbreeding has been reported in plaice, *Pleuronectes platessa* (Hoarau *et al.*, 2005), a close relative of dab, the genetic data obtained from dab were not suggestive of inbreeding in any case (all samples did not depart significantly from Hardy-Weinberg equilibrium). Even if the number of fish contributing to the next generation (effective population size) is orders of magnitude smaller than census size(Hauser *et al.*, 2002), the abundance of dab (Rijnsdorp *et al.*, 1992) would make the possibility of inbreeding (unless behaviourally driven) remote. Moreover, several of the loci employed were highly diverse (Chapter 6), and consequently, levels of heterozygosity were also correspondingly high which might confound potential signals(David, 1998). Perhaps a single locus heterozygosity analysis, to take into account differences in allele diversity, might provide a better resolution between individual locus genetic diversity and biomarkers of pollution exposure (Goudet & Keller, 2002). Additionally, although the inbreeding measure incorporating molecular distance, d^2, (Coulson *et al.*, 1998), has been refuted as relevant in most hypothetical cases (Tsitrone *et al.*, 2001), Goudet & Keller (2002) suggest that perhaps in what they consider very rare situations where recent admixture of large subpopulations had occurred, the differential fitness of the outcrossed and inbred individuals might be detectable with d^2. Therefore future analysis should be aimed at exploring the relationship between biomarkers of pollution (such as liver nodules) and single locus heterozygosity, with and without molecular distance, of those samples with evidence of recent admixture and population changes (LyB and InC).

Overall, although allozymes are better suited for studies of heterozygosity-fitness-correlations in polluted environments (Larno *et al.*, 2001; Marchand *et al.*, 2003; Peles *et al.*, 2003; Maes *et al.*, 2005), their lower polymorphism limits their value as discriminators of population of origin, a limitation that outweighs their application to biomonitoring programmes. Therefore, if a choice between the markers has to be made, microsatellites emerge as a more appropriate population genetic marker for biomonitoring programmes.

8.5.3 Familial propensity to tumour development:

There was no increased relatedness among fish suffering liver nodules, providing no evidence for family effects on the incidence of grossly visible liver nodules. The only case with positive relatedness among dab with liver nodules was in StB07, but in this case most fish were

allocated to Liv07, thus indicating the increased relatedness was probably an effect of the immigration rather than a single local "family" in StB07 being particularly prone to liver nodules. On the contrary, an interesting observation was that most groups composed of individuals afflicted with liver nodules were less related to each other than their healthy counterparts. The pattern was significant in one case (InC07), and nearly significant in two others (NeD08, InF08). These cases were associated with low probability of the individuals belonging to the same population as the rest of the sample and suffered from diseases not common in the rest of the fish collected in the same sample. Such a pattern could be either created by a genome x environment mismatch hypothesis (Garcia de Leaniz *et al.*, 2007) in which migrant individuals are poor performers in the novel environment and rapidly succumb to locally prevalent diseases; or alternatively, these individuals represent the lingering remnants of previously dominant local fish. In order to answer these questions a combination of genetic profiling and otoliths aging and otolith microchemistry (Thorrold *et al.*, 2001; Miller *et al.*, 2005) would be powerful at discerning between immigrants, either as early juveniles or recently as adults, and generational genetic changes, where both adult and juveniles show the same otolith microchemistry (due to the prevailing water chemistry) but different genetic signals.

8.6 Conclusion and Recommendations:

The incorporation of population genetics data into biomonitoring programmes has provided valuable insights into the aetiology of the observed prevalence of diseases, and more particularly liver nodules, around England and Wales.

Several cases of potential diseased migrants were revealed. The importance of these movement patterns is paramount in the correct interpretation of biomarker data, as some locations (e.g. StB in 2007) may be considered erroneously over-exposed to carcinogens. Conversely, some samples with differing biomarker incidence compared to other years in the same location were associated with genetic changes (e.g. InF06 and InC07), suggesting that their biomarker profile might not represent local exposure to pollutants. There were no temporally stable patterns of movement, which prevented the formulation of general trends of disease incidence due to migration.

No obvious relationship between multilocus heterozygosity as measured by microsatellite markers and disease profiles or proxies of fitness (age standardised weight and length) was found. Inbreeding in such a common marine fish is unlikely, therefore reducing the possibility of inbreeding-associated heterozygosity outperformance.

Genotyping of individuals included in biomonitoring programmes or at least in those stations of most concern, should be introduced and ideally carried out routinely to achieve an improved understanding of patterns of movement, biomarker incidence, and potential selection. Increasing the number of loci screened would improve the rejection power of the assignment test, giving more confidence to the results obtained. Therefore, it becomes paramount to develop cheaper methods of individual genotyping.

Particularly in dab, genotyping of samples covering the whole species range would increase markedly the power of assignment exercises and would help explain better the disease-genetic relationships between locations and years. Finally, genotyping of archived dab tissues of Cardigan Bay samples prior to the widespread emergence of carcinogenic processes would help interpret the temporal trends of increased disease incidence in the late 1990's.

Chapter 9: General Discussion

9.1 Aims and themes of the thesis:

Chemical compounds derived from human activities can have conspicuous and significant effects on marine organisms. Such degradation restricts the resources and benefits man draws from the marine environment. It is in mankind's interest to protect and preserve the natural resources of such an essential feature of planet. It therefore becomes important to develop strategies that can monitor marine environmental health using reliable and sensitive indicators, as well as applying appropriate tools to minimise detrimental consequences and facilitate recovery.

Below, I provide a critique of the design of the study described in this thesis, with emphasis on constraints imposed by the available biological system and markers employed. This is followed by comments on both, biomonitoring programmes, and the use of molecular markers to detect differentiation and selection. Finally, some potential ideas for future work will be proposed.

Understanding the effects of pollutants in particular environments is fundamental to generating the policy and regulations that aim to manage and control pollutants. Controlled laboratory experiments are well suited to understanding the effect of one chemical compound in a particular set of conditions, but they cannot emulate the complexities of the reality experienced by wild organisms. Therefore, studying directly the effects of man's largest experiment, the modification of Earth's environment, is vital for the continued existence of humankind. Biomonitoring programmes aim to evaluate the status of natural ecosystems, which combined with known anthropogenic disturbance, help understand the complex interaction between wildlife and human activities.

Biomonitoring programmes measure changes correlated with pollution exposure in bioindicator species, and draw inferences of pollutant bioavailability from the observed changes. However, individual genetic variability can translate into different responses to pollutants. Such differences can be decisive in the survival of an individual. If the differential survival is indeed due to particular gene combinations and their associated traits, then selection may lead to adaptation of the local population to pollutants. Local adaptation is only possible if the effects of selection in one generation are not diluted by an influx of immigrants. Therefore, a group of individuals may have the *potential* for local adaptation if it

is relatively isolated from other such groups of individuals. The relevance of local adaptation for biomonitoring programmes is that, if the measured biomarker of exposure is detrimental for the survival of the individual, then selection over generations in the local population may show reduced values for the biomarker measured (i.e. the population is pollution-adapted). Such findings may lead to the erroneous conclusion that pollution effects are receding. On the other hand, if the number of immigrants is sufficiently high to prevent local adaptation, individuals may not indicate proximate levels of pollution exposure.

Given the key role that movement and pollution selection can have on the reliability of biomonitoring results, the current thesis aimed to evaluate the extent and patterns of genetic structuring (driven by gene flow, selection and genetic drift) in flatfish species employed in the UK as bioindicators of pollution exposure. To that aim, a large number of genetic markers, both neutral and potentially adaptive, were developed for two species: dab, *Limanda limanda*, and flounder, *Platichthys flesus*.

Unfortunately, due to lack of substantial number of samples, time restrictions, and competition with other projects also working on population genetic structure of flounder, no further analysis on flounder was performed after the development of microsatellites. Nevertheless, in view of the documented record and continuing interest in flounder and its response to contaminants (Kirby *et al.*, 1999; Stentiford *et al.*, 2003; Kirby *et al.*, 2006), the new microsatellites will facilitate additional work on the extent and dynamics of population structuring in this species.

A temporally replicated and comprehensive sampling regime of dab around Great Britain was accomplished. Over 3000 individual fish were genotyped for 16 microsatellite loci of varying levels of allelic diversity. Highly significant and temporally stable differentiation was detected between North Sea and Irish Sea fish, indicating the existence of at least two genetically distinct populations of dab.

9.2 Critical evaluation of the project:

9.2.1 Movement of dab:

Individual movement patterns are of paramount importance for biomonitoring programmes. In the current study several instances of potential movement of large groups of fish, as well as individual fish were suggested by the combination of genetic and biomarker datasets. No general trends were detected across years, and thus no routine migrations could be inferred.

Such findings are highly relevant in the interpretation of biomarker information, as the individual-estimated exposure to pollutants could be very different from historical exposure.

However, several issues undermine the interpretation of the movement results: First of all, although most of the diseased migrant fish were assigned to populations in the same basin in which they were collected, the level of genetic differentiation among samples within basins was low, and thus confidence in the assignment was correspondingly low. Second, most of the species range remained unsampled, thus migration of individuals from unsampled populations could impact negatively on the assessment exercise. And finally, the lack of age data proved problematic when explaining the differences observed between diseased and non-diseased fish collected in the same location, as without age it was not possible to distinguish between recent migration of the diseased fish and local older fish belonging to a genetically different cohort from non-diseased fish.

However, the potential for diseased fish to be genetically different from non-diseased fish was demonstrated, as was the possibility that the occasional occurrence of liver injuries in dab in some sites derived from recent migrants. Identifying migrants and source populations would be of great advantage in biomonitoring programmes of mobile species, as not only will the interpretation of biomarker data be more precise, but a wider perspective of the dispersive effects of localised pollution will be gained.

9.2.2 Selection imposed by pollution:

The existence of genetic structure indicates the possibility of independent evolution of the different gene pools of the species (Carvalho, 1993; Conover *et al.*, 2006), and several biological traits are indeed different between North Sea and Irish Sea fish. Therefore, dab from the North and Irish Seas may have been subjected to different natural selection pressures (and independent genetic drift) and may well have differences in their biological response. However, it does not necessarily follow that these groups have the potential for pollution-induced adaptations, as pollution is unlikely to exert selection on whole sea basins. The evidence for structuring within sea basin was much weaker, although still significant in some years. North East Dogger in particular was significantly different to other sites in the North Sea for several years, which together with the high metal concentration and the potential for dab to settle offshore, may open the possibility for local adaptation of dab in this particular site. Several methods were employed aimed at detecting genetic differences between diseased and non-diseased fish at Dogger Bank:

The results of the gene-linked microsatellites were inconclusive. However, several aspects of the experimental design may have reduced our power to detect selection. First of all, the use of only taxonomically conserved genes may have been an inappropriate choice, as if the genes are present in both dab and other divergent taxa, the genes are probably tightly functionally restricted thus limiting the variability for selection to act upon. Secondly, the age of the fish used was unknown when the experiments were performed, and thus the diseased and non-diseased phenotypes may have consisted of fish of different age classes, thereby compromising the comparison. The test would have been more meaningful if only old fish (i.e. >5 years) with, and without disease were used. Finally, the gene-linked microsatellites should have been genotyped in more single individuals to gain statistical power in those cases where reduced polymorphism was detected. In retrospect, the development of the gene-linked markers should have been performed in parallel with the construction of the neutral microsatellite library and incorporated into the routine genotyping of dab from the start. Despite not fulfilling the original aim of detecting genes under selection and differential survival, the project demonstrated, not only the possibility of developing polymorphic microsatellites very economically from EST libraries, but also that, for these, polymorphism can be "estimated" beforehand by comparing with other homologous sequences isolated in other species and individuals. Of the 35 primer pairs tested, 15 yielded polymorphic microsatellite loci in dab. The fact that these loci are tightly linked to genes which are up or down regulated after exposure to pollution, might render the set of 15 microsatellites attractive for studies of pollution effects on other flatfish, or even other fish species given the existence of conserved genes and regulatory pathways across taxa. And finally, the DNA pooling and selective sweep approach, although not ideally suited for marine organisms, remained a fast and economic method to detect polymorphism and variance in that polymorphism among groups of individuals. The method could be particularly useful when studying strong selective pressures and gradients, such as those imposed by salinity in the North Sea-Baltic confluence (Andersson *et al.*, 1981; Nielsen *et al.*, 2004; Hemmer-Hansen *et al.*, 2007b; Larsen *et al.*, 2007; Limborg *et al.*, 2009), or when studying the effects on the genetic composition of historical pollution burdens such as in the Mersey Estuary or Puget Sound (Jones, 2006; Myers *et al.*, 2008).

A second attempt was made to test whether diseased fish belonged to the same genetic population as the non-diseased fish collected in the same location (regardless of which one represented the "local" population). Assignment tests can be interpreted in two ways: first, to find the source population of an individual; and secondly, whether sampling sites can be

rejected as potential source populations. Of the 61 fish encountered with liver nodules, only three could be significantly rejected as belonging to capture sites. Given the limited genetic differentiation between neighbouring sampling sites, such results are unsurprising. The relatedness test was more powerful when more than one individual per sample suffered from liver nodules, and evidence was found that some diseased fish were not a random subset of fish collected in the location. Such findings are intriguing and demand further study. As mentioned earlier, it is not known whether these fish represent true migrants or are genetically different local fish (either by age or susceptibility), as both hypotheses imply similar genetic signatures (diseased fish are genetically different from non-diseased fish). If alternative methods could be employed to provide an independent estimation of origin (Smith *et al.*, 2002; Selkoe *et al.*, 2008), the issue would be better resolved (Section 9.3.3). Regardless of whether the differences were created by selection, cohort differences or migration, it was still evident that in some locations diseased fish did not belong to the same breeding unit as proximate non-diseased fish.

A third and final attempt of detecting differences between diseased and non-diseased fish was evaluated: the relationship between disease and heterozygosity. No significant relationships were, however, found. There were several problems which limited the potential to detect selection in the current study; some were analytical (already mentioned); while others were methodological or inherent qualities of the system studied.

First, there were problems with the choice of samples. In most cases the tissue samples received for genotyping belonged to those fish only assessed for external diseases and grossly visible liver nodules, thereby increasing the sample size (from 50 to over a 100), though limiting the potential to correlate the genetic information to more detailed biomarker data. Genotyping the 50 fish fully evaluated for biomarkers plus another extra 50 evaluated only for external diseases, would have been a more appropriate genotyping regime. In doing so, the complete biomarker data would still be available for half of the individuals, while meeting the need for large sample sizes to detect subtle marine population structure (Waples, 1998).

The system studied was also not ideally suited for studies on pollution-induced selection. The localised occurrence of pollution combined with a fish capable of regional movement imposes inherent uncertainty over the life-long exposure to pollutants. However, such a scenario is of course typical of many marine taxa, and studies need to take the consequences of such factors into consideration when generating sampling designs. Second, the low levels of the targeted biomarker (liver nodules) in current populations prevented statistical analysis

between diseased and non-diseased fish. Third, dab reach sexual maturity much earlier (age one or 2) than the age at which most biomarkers are detected, thereby contributing to the next generation before potential differential mortality. Finally, the mortality directly attributed to pollution is unknown, but given the generally low disease levels (maximum of 18% in StB07 though normally 5%-10% in the most polluted sites), the extent of pollution-induced selection in dab is difficult to predict. Nevertheless, dab larvae, in contrast to plaice or flounder, are not dependent on coastal shallow areas and estuaries to settle (Henderson, 1998), and thus are hypothetically able to directly settle in the same environment as their parents, which *ecologically* may enhance the potential for local adaptation. Furthermore, differences in overall-lifetime reproductive success between susceptible and tolerant fish might be a sufficiently strong selective pressure, especially since it is females in general who live longest (Deniel, 1990). Older females tend to produce more eggs with higher survival chances (Buckley *et al.*, 1991), and females are three times more likely to develop liver injuries than males of the same age (Koehler, 2004). The combination of stronger selective pressures on females, with larger reproductive output of more viable larvae by pollution-tolerant females, could lead to a selective advantage for individuals born from native older females in polluted sites such as Dogger Bank. Similar reasoning can be applied also to putative selection against immigrant genotypes, which is increasingly recognised as a key mechanism of population differentiation (Nosil *et al.*, 2005).

Pollution-induced selection is more easily studied in limited gene-flow systems such as lakes and rivers subject to industrial effluent (Theodorakis & Shugart, 1998; Theodorakis *et al.*, 2006; Bourret *et al.*, 2008), though, geographic-specific selection in marine fish is still possible under strong environmental pressures or gradients as evidenced by locally adapted populations of flounder (Nissling *et al.*, 2002; Hemmer-Hansen *et al.*, 2007a; 2007b; Larsen *et al.*, 2007; 2008), herring (Andersson *et al.*, 1981; Bekkevold *et al.*, 2005; Ruzzante *et al.*, 2006) and cod (Knutsen *et al.*, 2003; Pogson & Fevolden, 2003; Jorde *et al.*, 2007; Olsen *et al.*, 2008).

When aiming to study pollution-induced selection, semi-enclosed areas of the sea where diseases have a much higher prevalence, such as the 80% liver nodule incidence reported from Puget Sound in the 1980's (Myers *et al.*, 2008), would be an ideal study sites. Nonetheless, such high incidence of diseases is fortunately not recorded from British waters.

9.3 Biomonitoring programmes and genetics:

The work presented here highlighted the importance of evaluating patterns of connectivity and selection in biomonitoring. First of all, a quick recap over some key factors and characteristics that define ideal bioindicator species and biomarker responses. The former should live in close contact with pollution, be widespread and amenable to frequent sampling, and should represent the local conditions with respect to pollution exposure. The latter should have a reliable dose-response relationship with exposure, and not be affected by non-pollutant factors, such as season, size, sampling protocol or individual variability (Wu *et al.*, 2005).

Therefore, it is important to identify clearly the significance of connectivity and adaptation on such factors from a biomonitoring perspective: movement of individuals (not to be confused with what geneticists define as migration or gene flow) will mainly affect the use of bioindicator species, as a "good" biomarker will still represent the changing surroundings wherever the individual goes; on the other hand, adaptation will principally affect the use of a particular biomarker, as individual variability, and selection acting on it, may influence the biomarker response of an otherwise effective bioindicator. Molecular genetic tools can assist in understanding these processes in biomonitoring schemes.

9.3.1 Movement of bioindicators:

Sources of pollution are probably patchy in the natural environment; individual movement therefore introduces variability in the exposure to pollution and consequently into the biomarker being measured. As described previously, groups of individuals may be relatively isolated from other such groups, which, through random genetic drift, will differentiate genetically over generations creating subpopulations. Such subpopulations can be identified by genetic markers. Movement of individuals may be more widespread than the extent of the subpopulation, but effective reproduction with other subpopulation may be affected by either natal homing (Thorrold *et al.*, 2001) or selection (Nosil *et al.*, 2005). Therefore, it is possible to detect with genetic markers recent immigrants or passing individuals that do not belong to the local populations. In this thesis, potential movement among dab biomonitoring stations was uncovered with the use of genetic markers, raising uncertainty about the legitimacy of sampled individual dab as being representative of local levels of pollution. However, the general temporal stability in biomarker frequency among biomonitoring stations (Stentiford *et al.*, 2009) suggests that most dab actually represent local fish. Furthermore, the existence of very small but significant genetic differences among some

locations within sea basin suggests subpopulation cohesiveness giving more support to the biomonitoring data. Indeed, the benefits of including population genetic tools in biomonitoring programmes using mobile bioindicators are two-fold: first, the genetic discreteness of bioindicators among stations at the time of sampling can be interpreted as relative stability of the local population, if the pattern is replicated temporally. Second, rare migrants with non-concordant phenotypes or biomarker profile can be detected, removing some of the variability originally introduced by movement of individuals.

9.3.2 Biomarker response variability:

Organisms reproducing through sex are genetically variable, and such genetic variability, if coding, can translate into advantages for some individuals under certain environmental circumstances. Pollution, like any other environmental variable, can impose strong selective pressures. Indeed, pollutants have been shown to impact genetic diversity (allele richness and heterozygosity), both experimentally (Gardeström *et al.*, 2006; Nowak *et al.*, 2009), and in wild populations (Theodorakis & Shugart, 1997; Larno *et al.*, 2001; Peles *et al.*, 2003; Theodorakis *et al.*, 2006; Bourret *et al.*, 2008). As seen earlier, if the biomarker employed conveys any benefit or hindrance to individuals, the biomarker response may be under selection, which, over generations and isolation, may lead to a locally adapted population and modified response biomarker profile. Such processes will be of particular importance for those bioindicator species with reduced mobility of larval dispersal. Untersee & Pechenik (2007) compared the copper tolerance on juveniles of two closely related gastropods, one with a pelagic larvae, *Crepidula fornicata*, and another with a benthic larvae, *C. convexa*, from polluted and reference sites. The juveniles of the latter from the polluted site showed significantly less mortality after 96h exposure to copper than those from the reference site and the other species, suggesting that *C. convexa*, given its non-dispersive reproductive mode, had adapted to a high copper concentrations. Although the experiment was aimed at assessing a biomarker, the concept remains the same, that organisms with limited dispersal (whether larval or adult) may elicit different biomarker response depending on previous exposure. Again molecular genetic markers prove invaluable here in detecting the potential for local adaptation, when using neutral markers, or actual changes in pollution response, when using adaptive genetic markers.

The existence of two genetically different dab populations indicate that fish in each basin might evolve independently from each other. Therefore, it was suggested that biomonitoring programmes acknowledge the results of the genetic analysis and avoid the comparison

between North Sea and Irish Sea, as fish from either basin have the potential to respond differently to common environmental pressures. As mentioned earlier, there are several other biological traits which are also different between both populations, including biomarker tendency and age at first appearance of liver nodules (G. Stentiford – pers. communication). However, currently there is no direct evidence that the genetic difference at neutral markers translates into biomonitoring-relevant traits. Establishing such causality would require transplantation and laboratory common-garden experiments with fish from both sea basins. Some such experiments have been undertaken with flounder with interesting results. Kirby *et al.* (2006) sampled flounder from three different estuaries in the UK (Alde, Tyne and Mersey) and a farmed sample, and exposed the different fish to a vitellogenin inducer. They then measured the differences in biomarker response (vitellogenin level), and compared them among estuaries. All fish showed similarly high levels of biomarker response after six days, but after 10 days Mersey estuary flounder showed significantly lower induction on average than other fish samples. Most interestingly, fish from the Mersey and Tyne showed much more variability in response after 10 days exposure than fish from the reference sites (Alde and farmed), which all invariably expressed much higher vitellogenin levels. Such results may indicate that there are individuals in the Mersey with reduced response compared to reference estuaries. Flounder are much more likely than dab to be under strong pollution selective pressures as they inhabit estuaries for long parts of their lives where exposure levels may be higher.

An important evolutionary aspect of biomarker response is their heritability, as this will dictate whether adaptation is possible. Population of killifish, *Fundulus heteroclitus*, from contaminated sites show reduced EROD induction compared to non-exposed populations. First generation offspring inherit the low EROD activity when reared under controlled laboratory conditions, but not those from the second or third generation where the low EROD induction trait is apparently lost (Meyer *et al.*, 2002). The existence of genetic diversity at key pollution related genes, the role of such genetic diversity in individual pollution tolerance, and the strength of the selective pressures acting on bioindicator species should be evaluated.

9.3.3 Overall conclusion of the importance of evolutionary theory and genetic markers in biomonitoring:

Two key components of evolution, connectivity and selection, have important implications on the choice and interpretation of both bioindicator and biomarker. The variance in biomarker

introduced by the movement of mobile bioindicator species may initially suggest that sessile species are more suitable for assessment of local pollution exposure. However, the increased potential for selective forces on poorly dispersing species may locally modify the biomarker response. The effects of one may counterbalance the other, and as evaluated by Wu *et al.* (2005) no animal group (crustacean, mollusc nor fish) appear to offer improved biomonitoring potential. In an ideal world, all individuals of bioindicator species should respond equally to pollution exposure (i.e. show no variability or no heritability), and should represent local conditions. Detecting adaptive changes in natural populations is notoriously more difficult than detecting genetic structure, and correcting for individual movement is, thus, easier than correcting for variance in response.

The congruence of population of origin for most dab with liver nodules within a particular sea basin each year was astonishing (e.g. most fish with liver nodules collected in the Irish Sea originated in Liverpool Bay in 2007). Such findings demand further investigation as the implications are relevant to both the biology of dab and the evaluation of pollution bioavailability and its consequences. For example, in the Irish Sea, dab with liver nodules were mostly assigned to those areas with known high incidence of liver nodules, implying: first, that the pollution ranking of the location where the fish was collected should be lowered; second, that the pollution at the location to which the fish was assigned to, has a higher impact on dab than previously assumed; and third, that dab, and many other organisms, may be acting as a dispersive agent of pollutants from one area to others.

To conclude, the appreciation of evolutionary processes in biomonitoring is paramount for the correct evaluation of the biological effects of pollution, and genetic markers should be employed to correct for movement of bioindicators and potential adaptation processes in biomarkers.

9.4 Detecting differentiation and selection in marine fish using genetic markers:

Although differentiation and selection are two different processes they both have to potential to generate a genetic imprint on natural populations, and results of the latter cannot be understood without the consideration of the former. Detecting both microevolutionary forces in marine organisms is challenging. Many marine taxa combine large effective population size with dispersive life stages which may result in extensive connectivity and apparent lack of population structure. Such characteristics demand the use of the most sensitive markers and analytical techniques (Conover *et al.*, 2006; Hauser & Carvalho, 2008; Naish & Hard, 2008).

The structure found in dab contributes to the mounting number of studies challenging the paradigm that the sea is a homogeneous environment, indicating how marine fish with pelagic eggs and larvae can occur in temporally stable subdivided entities even without palpable environmental gradients such as salinity or temperature. Within the North East Atlantic there is evidence of structured populations, among others, of herring (Mariani *et al.*, 2005; Ruzzante *et al.*, 2006), sprat (Limborg *et al.*, 2009), hake (Lundy *et al.*, 1999; Castillo *et al.*, 2004), and cod (Hutchinson *et al.*, 2001). Structure has been found, even at very small geographical scales (Knutsen *et al.*, 2003; Jorde *et al.*, 2007). Gradually, scientists are accepting that marine fish can exist in localised populations subject to a mosaic of environmental pressures shaping the biological characteristics of individuals. Acknowledging that such individual variability may be linked to the environment where the individual resides should change the way many ecological aspects of marine organisms are understood and examined (Bembo *et al.*, 1996; Nielsen *et al.*, 2004; Hemmer-Hansen *et al.*, 2007a; Larsen *et al.*, 2007; Larsen *et al.*, 2008).

The relatively large number of microsatellite loci employed, combined with a temporally replicated sampling strategy proved powerful at detecting the structure of dab around the British Isles. The length of the study, four years, offered a rare view of the complexity of the genetic structure of marine fish. Some locations were genetically indistinguishable every year, a reflection of their putatively stable demographics, whereas in other cases, the genetic signal suggested either small populations exhibiting stochastic fluctuations or immigration from other areas.

The differing estimates of population structure and extensive data set available in the present study provided an opportunity for a thorough analysis of the different estimators of differentiation available, and an assessment of the relevance, if any, of allelic diversity *per se* on such estimates (Chapter 6). The overall conclusion indicated that substantial amounts of *consistent* structuring, that is, allele frequencies that were *consistently* different for several loci between sampling locations, remained unreported by the traditional estimators (e.g. loci *DAC4-64*, *DAC1-90*, and *DAC4-40*). The temporal and spatial consistency of the allele frequency differences between North Sea and Irish Sea provided unambiguous evidence of the importance of such heterogeneity.

The consequences of the limitations imposed by high allelic diversity on the estimates of differentiation are far from trivial: it demands a reconsideration of the way the choice of

molecular markers is made, the analysis of marine population structure is designed, and the assessment of the existence of selection and adaptation is conducted.

9.4.1 Choice of markers and analysis implications:

Various perceived constraints in the use of microsatellite markers, including limitations resulting from their highly variable levels of polymorphism (Olsen *et al.*, 2004), uncertain mutation models (Balloux & Lugon-Moulin, 2002) and issues of cross-calibration among laboratories (LaHood *et al.*, 2002) has stimulated a search for additional marker systems, especially the use of single nucleotide polymorphisms (SNPs). SNPs are variations in individual bases (A, C, G, and T) of a particular DNA sequence. They are normally biallelic, and can be homozygous for either allele or heterozygous. Several characteristics revolving around their low mutation rate have been highlighted as beneficial at inferring population history: low variation within populations, the ease of comparison between loci based on their binary nature, ease of interpretation, and presumed low levels of homoplasy (Brumfield *et al.*, 2003).

9.4.1.1 Effects of markers on the power to detect differentiation:

Genetic structuring detected with microsatellites that typically display high mutation rates has been viewed as unrepresentative of the general background level of variability (Brumfield *et al.*, 2003). However, such an assertion depends in part on the aim of the study. If the aim is to delineate groups of individuals whose "background" variability is strongly divergent (i.e. different species), then the low mutation rates of SNPs will give a more robust and conservative indication of reproductive isolation. Conversely, the low variability (low discerning power) combined with their low mutation rate (slow rate of change in allelic states) will weaken their power at detecting very recent and subtle structuring. As reviewed by Hauser & Carvalho, (2008) "very recent and subtle structuring" does not necessarily imply large numbers of effective migrants, especially in the context of large effective population sizes. If populations show temporally and spatially stable differences in allele frequencies at high diversity loci, it implies that those populations are not exchanging many individuals, confirming differentiation as suggested by Θ'_{WC}, and not as suggested by Θ_{WC}.

Rosenberg *et al.* (2003) established that microsatellites (particularly dinucleotides repeats) were five to eight times more informative than SNPs at detecting structure and ancestry in humans, and it is widely recognised that many more SNPs than microsatellites are needed to achieve the same level of statistical power measured as Θ_{WC} (Brumfield *et al.*, 2003; Seddon *et al.*, 2005; Ryynänen *et al.*, 2007; Schopen *et al.*, 2008). However, the low power of Θ_{WC}

compared to Θ'_{WC} to detect structure at high levels of heterozygosity may amplify the differences between both marker types. Payseur & Jing (2009) very recently published an excellent comparison of SNPs and microsatellites: Over 700 highly informative microsatellites (i.e. with alleles non-randomly distributed across populations) were analysed together with all SNPs located within 10,000 bases from each microsatellite. Therefore, microsatellites and SNPs were linked and should have gone through the same demographic and recombination events, that is, they should carry the same information regarding population structure. If so, the actual importance of mutational processes in our ability to of detect structure could be measured directly. Not surprisingly, differentiation measured as Θ_{WC} was higher for SNPs than for microsatellites, but when measures corrected for heterozygosity (G'_{ST}) and the "informativeness for assignment" (I_n) statistic (Rosenberg *et al.*, 2003) were employed, microsatellites provided much higher power for detecting population structure than neighbouring SNPs which had undergone the same demographic and recombination processes. Power to detect structure was positively correlated with the level of heterozygosity, as in the case of dab. Multi-SNP, haplotypes where several SNPs are inherited together, were suggested as having the potential to be as informative as microsatellites in human populations, though they concluded that in more recently diverged populations microsatellites would still outperform even multi-SNP haplotypes as markers of population structure.

The variance in mutation rates across microsatellite loci, which is so often considered a negative characteristic, might indeed be a source of rich phylogenetic information. Microsatellites should be regarded as single evolutionary entities, and averaging among them restricts the understanding of the distribution of allelic states. Different microsatellites have probably emerged at different evolutionary points in time, as some are polymorphic among widely divergent species (Primmer *et al.*, 1996), and thus will carry signatures of different past events. Furthermore, different repeat motifs mutate at different rates, and thus dinucleotides are expected to provide enhanced resolution in very recently diverged clades (Rosenberg *et al.*, 2003), while trinucleotides and tetranucleotides have similar detection power as SNPs (Payseur & Jing, 2009).

Overall, dinucleotide microsatellite markers are the most effective markers to detect population structure at low levels of differentiation, though traditional estimates of differentiation (Θ_{WC}) should not be employed to evaluate such differentiation.

9.4.1.2 Effect of marker on the power to detect selection:

Some SNPs are directly embedded within coding genes, which opens the possibility for selection to act on particular allele or heterozygosity states. Thus SNPs remain the choice marker when the selected point of mutation or target gene is known (Pogson & Fevolden, 2003). Furthermore, dense arrays of SNPs can provide enormous amounts of evolutionary significant information (Baird *et al.*, 2008). However, such SNP arrays are mostly unavailable for non-model species, and given that SNPs are rather uninformative unless in known selected genes or in large numbers, microsatellites remain the choice marker for detection of selection in non-model species. Nevertheless, the situation is changing as the detailed sequence or EST data on the genome of non-model species becomes more widely available (Pogson & Fevolden, 2003).

One way to evaluate signals of selection is to screen numerous genetic markers and compare their differentiation values with their heterozygosity (Beaumont & Nichols, 1996; Beaumont & Balding, 2004). Such a protocol is known as the *outlier loci* method and has become quite common. However, the correction for heterozygosity will most likely disrupt the relationship; indeed, the extent to which *outlier loci* can still be detected after correction for heterozygosity needs to be re-examined.

Alternatively, selection could be detected as loci who do not share the same *inertia* as the rest, that is, the differentiation follows a different pathway in a lineage of individuals which might then be detected spatially. If loci are putatively neutral they should only carry information of demographic and recombination events (an internal mutation rate), assuming that most of the genome is not affected by selection then, most of the scored loci should show concordant trends (albeit not concordant magnitude of differentiation, as this is highly dependent on allelic diversity). On the other hand if a locus is under selection, the signal provided by such loci, and others genomically linked, will be discordant to neutral loci or loci under stabilising selection. Exceptions may occur in extreme selection scenarios that lead to bottlenecks, in which selected and neutral loci will all show the same trends of inertia, but in these cases, signals of selective sweeps and bottlenecks are easier to detect (Cornuet & Luikart, 1996; Teschke *et al.*, 2008). The possibility to detect selection in the context of molecular markers displaying different properties should be explored with both simulations and empirical evaluation. However, to statistically test such assumptions one would need many loci, though this is also the case with the outlier loci method, and statistical problems regarding significance (5% false positives at 0.05 significance) (Vasemägi & Primmer, 2005) could still emerge.

9.4.2 Sampling strategy in marine organisms:

The emergence of geo-referenced individual-based analysis packages such as BARRIER (Manni & Guérard, 2004), GENELAND (Guillot *et al.*, 2005), BAPS2 (Corander *et al.*, 2004), and ADEGENET (Jombart, 2008) among others, is transforming the way studies in population genetics are designed and analysed (Manel *et al.*, 2003; Hansen & Hemmer-Hansen, 2007; Selkoe *et al.*, 2008). In the case of dab, these programmes proved invaluable for understanding the spatial distribution, connectivity and temporal aspects of populations.

Although sample data can be analysed, these packages have mathematical algorithms designed with individuals as the core unit of analysis, and can extract patterns of cohesiveness and differentiation based on such individual data. Thereby, it is suggested that sampling schemes should reflect these developments in statistical analysis. Indeed, collecting individuals across long stretches of the sea might provide very detailed information of the exact location of barriers to gene flow. Such sampling would enable researchers to delineate population boundaries of populations with greater accuracy, with consequent advantages for studies investigating selection: the actual geographical extent of a gene pool can be delineated in relation to a coherent and collective estimate of key environmental variation. In the case of dab, such a sampling strategy, which is inherent in the biomonitoring protocol, might have facilitated the identification of the putative location of the genetic barrier between the North Sea and Irish Sea populations. Furthermore, the drive to have large sample sizes was at the cost of some biomarker information in some samples (larger samples were screened for external diseases and liver nodules than for full biomarkers).

However, within-sample information such as HWE, linkage disequilibrium, and allele frequencies was of paramount importance to detect scoring errors because these produced deviations from those statistics. Smaller sample sizes would not allow checking the genotypic data as thoroughly and in the case of low differentiation systems: slight errors can have important impacts on estimates of differentiation, migration and other population parameters (DeWoody *et al.*, 2006). Furthermore, some biological systems are structured in such ways (i.e. isolation by distance) where there might be problems associated with estimating the number of populations from random individual sampling (Schwartz & McKelvey, 2009).

On the whole, the new geo-referenced programmes offer important advances on analysis of genetic data and improve understanding of patterns of connectivity and barriers to gene flow, but aiming for smaller sample sizes could potentially restrict the power to detect genotyping

errors. The balance between more detailed geographic coverage and minor increase in genotyping errors will depend on the biological system studied and the molecular marker employed.

9.4.3 Complementary tools to genetic markers:

Multidisciplinary assessments combining several types of markers can be powerful in detecting population structure (Selkoe *et al.*, 2008). In the case of dab, the biomarkers of pollution and genetic markers combined synergistically to highlight cases of individual migrants and population changes. The otoliths of dab in several samples were used to age the fish, which provided important insights into the relationship between age and many biomarkers of pollution. However, in several occasions the confounding effects of migration, possible generational genetic change, and selection remained unresolved. Fish age data, as read from otolith of the remaining samples, would have helped interpret some of these observations. Nevertheless, otoliths have other important attributes that should be considered in studies of population genetics, and particularly at distinguishing between selection, generational change, and migration.

Otoliths are formed throughout the life of an individual fish, a property that allows estimating the age of individuals. As they are produced, new layers of calcium carbonate capture the signal of trace elements present in the surrounding water. Therefore it is possible to cut thin sections of the otolith, analyse the microchemistry of a layer, and infer where the fish was when the layer was formed (Thorrold *et al.*, 2001), provided that sufficiently marked and consistent spatial or temporal heterogeneity exists. Otoliths may also vary in shape between fishery stocks, and thus shape has been used to delineate even between local stocks of cod, *Gadus morhua,* (Cardinale *et al.*, 2004) and orange roughy, *Hoplostethus atlanticus* (Smith *et al.*, 2002).

Bradbury *et al.* (2008) studied rainbow smelt, *Osmerus mordax*, among several estuaries, and found that adults were genetically different among estuaries, while recruiting juveniles were not. Such results could be the product of either natal homing, larvae are able to disperse but later return to their natal estuary to spawn, or alternatively, differential selection between local and immigrant juveniles resulting in the differentiated adult populations. The dilemma was resolved by using otolith microchemistry, which suggested the adults had not left the estuaries, and therefore concluding that selection against immigrant juveniles was the causative factor behind the differentiation among adult populations.

Otoliths have been used in plaice, *Pleuronectes platessa*, within the Irish Sea not only to detect population structure and patterns of movement, but also past history of pollutant exposure (Geffen *et al.*, 2003). Their utility, even at such a small scale as the Irish Sea, suggests they would be informative to differentiate between fish moving within sea basin. However, this in not to say that otoliths could replace genetic markers as assignment method in flatfish biomonitoring programmes: otoliths cannot provide information on selection processes. Integrated data sets that employ both data from otoliths and genetic markers to examine population structure and gene flow offer a particularly potent approach.

9.5 Suggestions for future work:

Apart from the independent evolutionary potential of the dab populations employed as bioindicators in different areas, the importance of patterns of movement, as revealed by genetic markers, in the interpretation of biomonitoring data was highlighted by the current study. Therefore, the use of genetic markers to correctly assign the source population of the individual is recommended on a routine basis. More affordable and reliable methods of genotyping individuals should be developed, so that biomonitoring and other end-beneficiaries of genetic tools, such as the forensic evaluation of fish products (Ogden, 2008), can routinely genotype large numbers of individuals. The age of fish played a pivotal role when trying to disentangle the genetic and disease relationships. Therefore, age should be a key element in studies comparing biomarker incidence and age. Reading otoliths is a labour intensive process, so new methods to age fish in large quantities should be developed. The potential use of otolith microchemistry to distinguish between the different sampling locations should be explored, as they would provide a useful control for distinguishing between migration and the effects of selection.

The long-term temporal stability of the dab structure should be explored by studying archived samples of fish. Exploratory work on wax-embedded liver tissues from previous histopathology exercises (personal observation) indicated that DNA extracted was of sufficient quality to amplify short microsatellites. Use of such resources should be further explored, especially with Cardigan and Liverpool Bay dab populations, where temporal changes in liver nodule frequency have been reported.

Cardigan Bay and Dogger Bank offer attractive scenarios to study both, the role of pollution-induced selection (anthropogenic effects), and potential differential sensitivity of populations to pollution stress (natural variation present regardless of pollution). Genetic changes recorded during the four years, combined with changes in disease incidence, generate a

complex pattern worthy of further investigation. The potential for differential biomarker responses between Irish Sea and North Sea dab (or even at a regional level between Cardigan Bay/Red Wharf Bay and Dogger Bank/Amble) should be explored. Studying the population genetic structure of flounder, both with neutral and potentially adaptive genetic markers, could yield interesting results.

The performance and behaviour of the different estimators of differentiation: traditional, corrected and D_{est}, should be explored with simulations to understand their variance around the real parameter to be estimated. The effects of loci with differing allele diversities and null alleles should be explored to assess their reliability. Finally, the possibility of detecting selection after correction for heterozygosity with the use of differentiation should be re-examined.

9.6 Summary of the thesis:

A case is made to enhance the integration of population genetic data into biomonitoring programmes, and as such it is worthwhile summarising the various ways that such integration may be of value. Genetic markers may yield information on various processes, including: (i) The extent of migration and gene flow, which can provide indirect information on distinguishing resident and migrant individuals; (ii) The assignment of individuals to putative source populations, which not only allow for a more effective examination of spatially variable pollution stressors, but importantly also impacts on the assessment and dynamics of population recruitment; (iii) The detection of selection and existence of locally adapted populations, either as the potential for such differentiation using neutral markers, or directly through analysis of adaptive genes. The advent of genomic approaches is especially relevant here; (iv) The estimation of demographics, including gene flow and effective population size: aspects of population structure that affect significantly the response to selection; (v) The extent of species that may be cryptic and inadvertently included in estimates of congeneric genetic or phenotypic variation.

In relation to the above generalities, several key points can be highlighted:

1) Biomonitoring programmes are invaluable to understand the effects of chemical pollutants in the natural environment. They provide important information on the bioavailability of pollutants and the biological and ecological implications of such pollutants. Biomarker responses are measured in bioindicator species, and results can be interpreted as exposure to pollution.

2) Movement, migration, isolation and selection can all have implications on the reliability of bioindicators and biomarkers to convey correct information of pollutant exposure.

3) As a case study, genetic markers, both neutral and potentially adaptive, were developed and tested on a commonly used bioindicator species in biomonitoring programmes in the North East Atlantic: the dab, *Limanda limanda*.

4) Two broad evolutionary units were identified in the North and Irish Seas, with further weak substructuring within sea basin, resulting in complex pattern of diverse potential for local adaption and pollution-induced selection.

5) Molecular tools proved invaluable in detecting patterns of movement of individuals and temporal changes in genetic composition of dab in sampling locations which correlated with changes in biomarker profile.

6) Although no pollution-induced selection was detected, new avenues of detecting adaptation were explored and key sampling and statistical factors for projects targeting pollution-induced selection were highlighted.

7) Empirical evaluation of different estimators of differentiation indicated that a considerable proportion of genetic differentiation was not being captured by traditional estimators of differentiation. Once corrected, a more accurate representation of the distribution of allelic states was obtained. Such findings may have implications in the choice of markers when organisms with weak structure are the target of study.

Today the world gave the most beautiful gift.

Long hours had I been working in the office,
when, finally, mentally exhausted,
but contented with the progress made,
I decided it was time to go home...
I was not tired, my eyes were still wide open,
and my mind firing like an atomic reaction,
I closed my dear laptop and prepared to go home,
when something suddenly struck me,
I crossed the huge window of the office,
and there it was...

An enormous and glorious sun
solemnly rising between the clouds.
The pale early morning blue
was the perfect background for those
oranges and pinks emanating from the yellow sphere.
The green of the trees either side of the city
blended in to form the perfect frame,
the sea, intimidated by such beauty,
tried to mimic the colours like a playful child.

I went out, on my way home,
as I rose the hill heading to my own little piece of the world,
I continued to admire this unique gift from the top of the hill
The dormant city laid by my feet, unbelievably quiet.
Only the birds, exploiting the void air, welcomed the new day,
and I felt like the sun had given me that gift only to me,
to remind me of the beauty of life

The air was fresh and I felt immensely happy that I was alive,
that I was there, and then.
If I died tomorrow, I'll be a happy dead man,
as I have been in heaven, heaven on earth....

Appendix:

Table A.1: Pairwise differentiation between male and female dab, *Limanda limanda*.

Pairwise differentiation (Θ_{WC}) between samples subdivided by sexes (females and males) and years (2005,2006,2007,2008). The lines within the squares represent the division between basins. The Θ_{WC} values have been colour coded relative to other values in the same table from yellow (very low) to orange (strong). Significance values are denoted with stars (~= near significance;*=below 0.05; **=below 0.01; ***=below 0.001), values significant after Bonferroni correction are in **bold**.

Dab females 2005

14 Loci	NeD05F	Amb05F		LyB05F		InC05F		StB05F	
NeD05F		0.000	0.478	0.009	0.019	**0.007**	0.000	0.018	0.024
Amb05F				0.011	0.018	0.001	0.232	0.019	0.018
LyB05F						**0.015**	0.001	0.016	0.078
InC05F								0.006	0.216
StB05F									

Dab males 2005

14 Loci	NeD05M	Amb05M		LyB05M		InC05M		StB05M	
NeD05M		0.007	0.034	0.004	0.060	**0.010**	0.000	**0.011**	0.001
Amb05M				0.001	0.339	0.005	0.115	0.004	0.175
LyB05M						0.003	0.108	0.004	0.117
InC05M								0.004	0.084
StB05M									

Table A.1: (cont.)

Dab females 2007

14 loci	NeD07F	Amb07F		OfF07F		InF07F		Rye07F		LyB07F		InC07F		RwB07F		Liv07F		StB07F	
NeD07F		-0.001	0.641	0.000	0.520	-0.001	0.714	0.002	0.002	0.000	0.373	0.002	0.194	**0.014**	0.000	**0.005**	0.000	**0.006**	0.000
Amb07F				-0.001	0.638	0.000	0.460	0.003	0.096	0.001	0.334	0.007	0.046	**0.020**	0.000	0.008	0.002	0.013	0.004
OfF07F						-0.002	0.864	0.001	0.198	-0.001	0.653	0.004	0.085	**0.020**	0.000	0.007	0.002	**0.014**	0.000
InF07F								0.000	0.396	0.002	0.170	0.005	0.056	**0.015**	0.000	0.006	0.002	**0.009**	0.000
Rye07F										0.000	0.349	0.000	0.496	**0.012**	0.000	**0.003**	0.000	0.003	0.028
LyB07F												-0.002	0.808	0.007	0.003	-0.001	0.875	0.004	0.067
InC07F														0.006	0.029	-0.001	0.763	0.002	0.295
RwB07F																0.003	0.031	0.004	0.062
Liv07F																		0.000	0.438

Dab males 2007

14 loci	NeD07M	Amb07M		OfF07M		InF07M		Rye07M		LyB07M		InC07M		RwB07M		Liv07M		StB07M	
NeD07M		-0.001	0.661	0.004	0.095	0.003	0.194	0.003	0.016	0.001	0.372	0.007	0.004	0.003	0.182	**0.007**	0.000	0.006	0.057
Amb07M				-0.002	0.651	-0.003	0.667	-0.004	0.992	-0.008	0.945	0.005	0.104	0.003	0.247	0.004	0.052	0.005	0.122
OfF07M						0.007	0.158	-0.001	0.597	0.001	0.417	0.013	0.005	0.014	0.014	**0.015**	0.001	0.016	0.016
InF07M								0.003	0.210	0.004	0.284	0.011	0.019	0.007	0.133	0.010	0.024	0.009	0.087
Rye07M										0.000	0.532	**0.009**	0.001	0.007	0.022	**0.006**	0.000	0.007	0.023
LyB07M												0.007	0.115	0.008	0.140	0.011	0.014	0.014	0.038
InC07M														-0.001	0.511	0.002	0.124	-0.001	0.592
RwB07M																-0.002	0.714	-0.002	0.634
Liv07M																		-0.002	0.667

Table A.1: (cont.)

Dab females 2008

FST	NeD08F		Amb08F		Off08F		Inf08F		Rye08F		LyB08F		InC08F		Liv08F		StB08F	
NeD08F			0.000	0.426	0.002	0.143	0.003	0.004	0.002	0.077	0.002	0.138	0.007	0.248	**0.007**	0.000	**0.010**	0.000
Amb08F					0.000	0.529	0.000	0.405	0.002	0.096	0.000	0.496	0.005	0.325	**0.006**	0.001	**0.010**	0.000
Off08F							0.002	0.130	0.002	0.207	0.006	0.012	0.010	0.217	0.010	0.003	**0.013**	0.000
Inf08F									0.001	0.271	0.002	0.119	0.004	0.369	**0.006**	0.000	**0.009**	0.000
Rye08F											0.001	0.332	0.008	0.250	**0.007**	0.000	**0.009**	0.000
LyB08F													0.002	0.366	0.005	0.015	0.007	0.004
InC08F															-0.008	0.648	-0.007	0.653
Liv08F																	-0.002	0.840

Dab males 2008

FST	NeD08M		Amb08M		Off08M		Inf08M		Rye08M		LyB08M		InC08M		Liv08M		StB08M	
NeD08M			0.000	0.476	-0.001	0.487	0.003	0.274	-0.002	0.557	0.011	0.150	0.009	0.068	0.000	0.436	0.001	0.411
Amb08M					0.002	0.091	0.007	0.011	0.000	0.499	0.011	0.020	**0.013**	0.000	**0.010**	0.000	**0.008**	0.001
Off08M							0.000	0.482	0.000	0.398	0.010	0.018	**0.008**	0.001	**0.011**	0.000	**0.006**	0.001
Inf08M									-0.002	0.764	0.009	0.064	**0.010**	0.001	0.012	0.002	0.004	0.053
Rye08M											0.008	0.090	0.006	0.050	0.008	0.004	0.005	0.020
LyB08M													0.005	0.190	0.010	0.057	0.011	0.019
InC08M															0.001	0.292	0.002	0.162
Liv08M																	0.004	0.031

Table A.2: Results of the Bottleneck analysis **(Next two pages)**

Results of the BOTTLENECK analysis for all dab samples by **location** (columns) and **years** (2005-08) with 14 loci. The probability of He being within the Heq 95% CI is reported for three different tests (**Sign**, **Standard differences** and **Wilcoxon**) under three different mutation models (IAM, TPM, and SMM). The Wilcoxon test also checks whether significance is due to heterozygote deficiency (expansion signal) or heterozygote excess (bottleneck signal). Significant values ($p<0.05$) are in red, those significant after bonferroni corretion are in red and **bold**. The results of the Standard differences test should be considered with care, as at least 20 loci are recomended for this test, thus they are shaded in grey.

Table A.2: (cont.)

				NeD	Amb	Off	Inf	Rye	LyB	Ces	Scb	Inc	RwB	Liv	StB	Dub	Noi	Woi
2005	Sign Test	IAM		0.464	0.237	0.096			0.533			0.544			0.408			
		TPM		0.069	0.345	0.158			0.159			0.175			0.338			
		SMM		0.004	0.062	0.023			0.005			0.003			0.067			
	Stand.	IAM		0.173	0.062	0.195			0.239			0.151			0.346			
		TPM		0.001	0.178	0.171			0.013			0.002			0.014			
		SMM		0.000	0.000	0.000			0.000			0.000			0.000			
	Wilcoxon	IAM	H def	0.852	0.932	0.923			0.768			0.866			0.687			
			H exc	0.163	0.077	0.086			0.251			0.148			0.335			
			H def/exc	0.326	0.153	0.173			0.502			0.296			0.670			
		TPM	H def	0.059	0.271	0.086			0.059			0.052			0.163			
			H exc	0.948	0.749	0.923			0.948			0.955			0.852			
			H def/exc	0.119	0.542	0.173			0.119			0.104			0.326			
		SMM	H def	0.001	0.021	0.003			0.000			0.002			0.015			
			H exc	0.999	0.982	0.998			1.000			0.998			0.988			
			H def/exc	0.002	0.042	0.005			0.001			0.004			0.030			
				NeD	Amb	Off	Inf	Rye	LyB	Ces	Scb	Inc	RwB	Liv	StB	Dub	Noi	Woi
2006	Sign Test	IAM		0.241	0.234	0.242	0.333	0.460	0.249	0.242	0.515	0.249	0.423	0.533	0.542	0.114	0.333	0.444
		TPM		0.171	0.386	0.325	0.313	0.070	0.343	0.182	0.160	0.313	0.358	0.130	0.071	0.062	0.069	0.166
		SMM		0.016	0.020	0.005	0.004	0.004	0.017	0.019	0.003	0.018	0.075	0.019	0.015	0.015	0.014	0.001
	Stand. Diff	IAM		0.031	0.032	0.288	0.144	0.085	0.074	0.025	0.135	0.139	0.066	0.209	0.158	0.114	0.187	0.341
		TPM		0.047	0.166	0.004	0.010	0.016	0.039	0.086	0.003	0.068	0.098	0.006	0.006	0.007	0.001	0.001
		SMM		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Wilcoxon	IAM	H def	0.955	0.955	0.805	0.821	0.923	0.923	0.961	0.805	0.879	0.923	0.821	0.805	0.914	0.787	0.729
			H exc	0.052	0.052	0.213	0.196	0.086	0.086	0.045	0.213	0.134	0.086	0.196	0.213	0.097	0.232	0.292
			H def/exc	0.104	0.104	0.426	0.391	0.173	0.173	0.091	0.426	0.268	0.173	0.391	0.426	0.194	0.463	0.583
		TPM	H def	0.292	0.271	0.179	0.108	0.121	0.213	0.271	0.121	0.213	0.292	0.121	0.134	0.108	0.052	0.059
			H exc	0.729	0.749	0.837	0.903	0.892	0.805	0.749	0.892	0.805	0.729	0.892	0.879	0.903	0.955	0.948
			H def/exc	0.583	0.542	0.358	0.217	0.241	0.426	0.542	0.241	0.426	0.583	0.241	0.268	0.217	0.104	0.119
		SMM	H def	0.003	0.008	0.002	0.001	0.001	0.015	0.010	0.002	0.003	0.029	0.001	0.002	0.002	0.002	0.001
			H exc	0.998	0.993	0.999	1.000	0.999	0.988	0.992	0.999	0.998	0.975	0.999	0.999	0.998	0.998	1.000
			H def/exc	0.005	0.017	0.003	0.001	0.002	0.030	0.020	0.003	0.005	0.058	0.003	0.003	0.004	0.004	0.001

TableA.2: (cont.)

				NeD	Amb	Off	Inf	Rye	LyB	Ces	Scb	Inc	RwB	Liv	StB	Dub	Noi	Woi
2007	Sign Test	IAM		0.471	0.445	0.217	0.559	0.460	0.219			0.238	0.416	0.492	0.430			
		TPM		0.061	0.019	0.023	0.060	0.019	0.187			0.020	0.021	0.016	0.022			
		SMM		0.019	0.018	0.021	0.004	0.004	0.075			0.018	0.005	0.004	0.022			
	Stand. Diff	IAM		0.043	0.116	0.024	0.296	0.047	0.024			0.091	0.132	0.135	0.227			
		TPM		0.000	0.000	0.000	0.000	0.000	0.000			0.000	0.000	0.000	0.000			
		SMM		0.000	0.000	0.000	0.000	0.000	0.000			0.000	0.000	0.000	0.000			
	Wilcoxon	IAM	H def	0.923	0.866	0.955	0.643	0.932	0.948			0.852	0.892	0.821	0.687			
			H exc	0.086	0.148	0.052	0.380	0.077	0.059			0.163	0.121	0.196	0.335			
			H def/exc	0.173	0.296	0.104	0.761	0.153	0.119			0.326	0.241	0.391	0.670			
		TPM	H def	0.008	0.003	0.045	0.012	0.007	0.039			0.029	0.003	0.003	0.018			
		90% smm	H exc	0.993	0.998	0.961	0.990	0.997	0.966			0.975	0.998	0.998	0.985			
		var=10	H def/exc	0.017	0.005	0.091	0.025	0.013	0.078			0.058	0.005	0.005	0.035			
		SMM	H def	0.002	0.003	0.025	0.001	0.001	0.034			0.003	0.001	0.001	0.007			
			H exc	0.998	0.998	0.979	0.999	0.999	0.971			0.998	0.999	0.999	0.997			
			H def/exc	0.004	0.005	0.049	0.003	0.002	0.068			0.005	0.002	0.002	0.013			
				NeD	Amb	Off	Inf	Rye	LyB	Ces	Scb	Inc	RwB	Liv	StB	Dub	Noi	Woi
2008	Sign Test	IAM		0.415	0.257	0.273	0.476	0.450	0.445			0.134		0.249	0.477			
		TPM		0.357	0.343	0.060	0.067	0.067	0.335			0.541		0.345	0.167			
		SMM		0.019	0.004	0.001	0.004	0.001	0.023			0.119		0.004	0.004			
	Stand. Diff	IAM		0.045	0.025	0.087	0.102	0.115	0.086			0.057		0.063	0.142			
		TPM		0.016	0.108	0.011	0.013	0.005	0.082			0.178		0.022	0.005			
		SMM		0.000	0.000	0.000	0.000	0.000	0.000			0.000		0.000	0.000			
	Wilcoxon	IAM	H def	0.932	0.971	0.892	0.866	0.892	0.923			0.953		0.948	0.821			
			H exc	0.077	0.034	0.121	0.148	0.121	0.086			0.055		0.059	0.196			
			H def/exc	0.153	0.068	0.241	0.296	0.241	0.173			0.110		0.119	0.391			
		TPM	H def	0.134	0.232	0.068	0.077	0.059	0.232			0.632		0.163	0.134			
			H exc	0.879	0.787	0.941	0.932	0.948	0.787			0.393		0.852	0.879			
			H def/exc	0.268	0.463	0.135	0.153	0.119	0.463			0.787		0.326	0.268			
		SMM	H def	0.008	0.007	0.001	0.001	0.001	0.002			0.064		0.001	0.001			
			H exc	0.993	0.997	1.000	0.999	1.000	0.998			0.945		0.999	0.999			
			H def/exc	0.017	0.013	0.001	0.003	0.001	0.004			0.127		0.002	0.002			

Table B.1: Correlations with individual inbreeding coefficient (I)

Correlation (r) between individual dab inbreeding coefficients, age, fitness proxies, sex and several biomarkers of pollution. Three samples were considered for the analysis: **NeD07**, **Rye07** and **Liv07**; ***IR***= Internal relatedness; ***hL***= average homozygosity by loci; **Age** = age of individual fish; **Lgt**= length; **Wgt**= weigth; **LgtST**= age-standardised length; **WgtST**= age-standardised weight; Values within the double edged box are those relating inbreeding coefficients to all other parameters. Associated significance values are depicted to the right of correlation values: **~** = p<0.1; ***** = p<0.05; ****** = p<0.01; ******* = p<0.001.

	NeD07						Rye07						Liv07					
	IR	***hL***	**Lgt**	**Wgt**	**Age**	**LgtST**	***IR***	***hL***	**Lgt**	**Wgt**	**Age**	**LgtST**	***IR***	***hL***	**Lgt**	**Wgt**	**Age**	**LgtST**
hL	0.964 ***						0.940 ***						0.955 ***					
Lgt	0.092	0.086					-0.079	-0.140					0.068	0.071				
Wgt	0.122	0.111	0.938 ***				-0.071	-0.146	0.971 ***				0.076	0.088	0.962 ***			
Age	0.109	0.100	0.845 ***	0.806 ***			-0.174	-0.196 *	0.782 ***	0.757 ***			0.125	0.121	0.829 ***	0.793 ***		
LgtST	-0.027	-0.018	-	0.388 ***	-0.002		0.153	0.090	-	0.487 ***	0.000		0.031	0.050	-	0.454 ***	-0.017	
WgtST	-0.034	-0.038	0.358 ***	-	-0.001	0.882 ***	0.134	0.065	0.421 ***	-	0.000	0.962 ***	0.011	0.033	0.414 ***	-	-0.020	0.953 ***

Table B. 2: Correlations with individual inbreeding coefficient (II)

Correlation (r) between individual dab inbreeding coefficients, age, sex and several biomarkers of pollution. Three samples were considered for the analysis: **NeD07**, **Rye07** and **Liv07**; ***IR***= Internal relatedness; ***hL***= average homozygosity by loci; **Age** = age of individual fish; **M**= male; **F**= female; **LY**=Lymphocystis; **U**= skin ulceration; **EP**= epidermal papilloma; **HYP**= Hyperpigmentation; **LN+**= presence or absence of liver nodules; **Cat1**= non-specific liver lesions; **Cat2**= non-neoplastic toxicopathic; **Cat3**= foci of cellular alteration; **Cat4** = Bening neoplasms; **Cat5** = Malignant neoplasms. Values within the double edged box are those relating inbreeding coefficients to all other parameters. Associated significance values are depicted to the right of correlation values: ~ = p<0.1; * = p<0.05; ** = p<0.01; *** = p<0.001.

NeD07	*IR*	*hL*	Age	M	F	LY	U	EP	HYP	LN+	Cat1	Cat2	Cat3	Cat4	Cat5
hL	0.964 ***														
Age	0.109	0.100													
M	-0.129 ~	-0.146 ~	-0.261 ***												
F	0.129 ~	0.147 ~	0.257 ***	-											
LY	-	-	-	-	-										
U	-0.054	-0.062	0.274 ***	-0.135 ~	0.136 ~	-									
EP	-	-	-	-	-	-	-								
HYP	-0.029	-0.048	0.488 ***	-0.094	0.097	-	0.206 **	-							
LN+	-0.001	-0.020	0.413 ***	-0.128 ~	0.129 ~	-	0.225 **	-	0.261 ***						
Cat1	0.207 **	0.198 **	0.284 ***	-0.101	0.100	-	0.090	-	0.146 ~	0.086					
Cat2	-0.006	-0.022	0.015	-0.037	0.038	-	-0.014	-	0.014	0.054	0.062				
Cat3	0.086	0.059	0.578 ***	-0.103	0.098	-	0.183 *	-	0.319 ***	0.280 ***	0.148 *	-0.034			
Cat4	0.024	0.033	0.520 ***	-0.054	0.055	-	0.205 **	-	0.278 ***	0.584 ***	0.109	-0.072	0.283 ***		
Cat5	0.071	0.017	0.107	-0.082	0.083 ~	-	0.222 **	-	0.099	0.380 ***	0.033	0.253 ***	0.140 ~	-0.038	
Paras	0.046	0.085	0.346 ***	-0.188 *	0.192 *	-	0.024	-	0.248 ***	0.101	0.072	0.037	0.196 **	0.083	0.061

Rye07

	IR	*hL*	Age	M	F	LY	U	EP	HYP	LN+	Cat1	Cat2	Cat3	Cat4	Cat5
hL	0.940 ***														
AGE	-0.174 ~	-0.196 *													
M	-0.220 *	-0.217 *	0.187 ~												
F	0.232 *	0.222 *	-0.180 ~	-1.000											
LY	-	-	-	-	-										
U	0.077	0.124	0.147	-0.080	0.081	-									
EP	-0.091	-0.058	0.271 **	0.118	-0.118	-	-0.009								
HYP	-0.027	0.011	0.401 ***	0.092	-0.091	-	-0.016	-0.016							
LN+	-	-	-	-	-	-	-	-	-						
Cat1	0.030	0.029	-0.035	-0.009	0.005	-	0.042	-0.223 *	0.074	-					
Cat2	0.148	0.159	-0.102	0.118	-0.118	-	-0.009	-0.009	-0.016	-	0.042				
Cat3	-0.041	-0.102	0.297 **	-0.061	0.063	-	-0.019	-0.019	0.265 **	-	0.086	-0.019			
Cat4	-	-	-	-	-	-	-	-	-	-	-	-	-		
Cat5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Paras	-0.074	-0.055	0.179 ~	0.058	-0.056	-	-0.028	-0.028	-0.049	-	0.047	-0.028	-0.057	-	-

Liv07

	IR	*hL*	Age	M	F	LY	U	EP	HYP	LN+	Cat1	Cat2	Cat3	Cat4	Cat5
hL	0.955 ***														
Age	0.125	0.121													
M	0.015	0.023	-0.340 ***												
F	-0.015	-0.023	0.340 ***	-											
LY	-	-	-	-	-										
U	0.096	0.088	0.184 *	-0.093	0.093	-									
EP	0.147 ~	0.178 *	0.141 ~	-0.090	0.090	-	0.102								
HYP	-0.003	0.007	0.116	-0.104	0.104	-	-0.044	-0.020							
LN+	0.098	0.084	0.141 ~	0.006	-0.006	-	0.172 *	-0.018	-0.020						
Cat1	-0.024	-0.015	0.447 ***	-0.183 *	0.183 **	-	0.002	0.066	0.076	0.066					
Cat2	-0.004	-0.032	0.118	-0.060	0.060	-	0.096	-0.025	-0.029	0.459 ***	0.094				
Cat3	0.045	0.059	0.415 ***	-0.128 ~	0.128 ~	-	0.099	-0.058	-0.068	0.062	0.178 **	0.002			
Cat4	0.059	0.066	0.312 ***	-0.030	0.030	-	0.111	-0.029	-0.034	0.392 ***	0.109	0.259 ***	0.127 ~		
Cat5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Paras	-0.019	-0.003	0.488 ***	-0.241 **	0.241 **	-	0.009	0.163 *	0.056	-0.009	0.252 ***	0.070	0.161 *	0.081	-

Bibliography

Abajian C (1994), [Online], Available: http://espressosoftware.com/pages/sputnik.jsp.

Abdullah MI, Royle LG (1972) Heavy metal content of some rivers and lakes in Wales, *Nature*, **238**, 329-330.

Abdullah MI, Royle LG, Morris AW (1972) Heavy metal concentration in coastal waters, *Nature*, **235**, 158-159.

Ahn Y-H, Shanmugam P, Lee J-H, Kang Y (2006) Application of satellite infrared data for mapping of thermal plume contamination in coastal ecosytem of Korea, *Marine Environmental Research*, **61**, 186-201.

Albert OT, Eliassen JE, Høines Å (1998) Flatfishes of Norwegian coasts and fjords, *Journal of Sea Research*, **40**, 153-171.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques, *Nucleic Acids Research*, **25**, 4692-4693.

Alonso-Amelot M, Castillo U, Smith B, DR L (1998) Excretion, through milk, of ptaquiloside in bracken-fed cows. A quantitative assessment, *Lait*, **78**, 413-423.

Amos W, Balmford A (2001) When does conservation genetics matter?, *Heredity*, **87**, 257-265.

Amos W, Worthington Wilmer JFK, Burg T, Croxall J, Bloch D, Coulson T (2001) The influence of parental relatedness of reproductive success, *Proceedings of the Royal Society of London Series B*, **268**, 2021-2027.

Andersson L, Ryman N, Rosenberg R, Ståhl G (1981) Genetic variability in Atlantic herring (*Clupea harengus harengus*): description of protein loci and population data, *Hereditas*, **95**, 69-78.

Andrews M, Rickard D (1980) Rehabilitation of the Inner Thames Estuary, *Marine Pollution Bulletin*, **11**, 327-332.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method, *BMC Bioinformatics*, **9**, Artcl. 323.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci?, *Molecular Ecology*, **15**, 4659-4665.

Ashcroft CJ, Sheffield E (1999) Rejection of *Pteridium aquilinum* subspecies <atlanticum(C.N. Page), *Botanical Journal of the Linnean Society*, **130**, 157-170.

Astanei I, Gosling E, Wilson J, Powell E (2005) Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas), *Molecular Ecology*, **14**, 1655-1666.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis AL, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers, *PLoS One*, **3**, e3376.

Baker RJ, Bickman A, Bondrkov M, Gaschak S, Matson C, Rodgers B, Wickliffer J, Chesser R (2001) Consequences of polluted environments on population structure: the bank vole (*Clethrionomys glareolus*) at Chernobyl, *Ecotoxicology*, **10**, 211-216.

Bakhsh AA (1982) *Population studies of the flatfish Limanda limanda (L.) in Anglesey waters*, Bangor: Ph.D. Thesis, Bangor University.

Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatellites can be misleading: an empirical and simulation study, *Evolution*, **54**, 1414-1422.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers, *Molecular Ecology*, **11**, 155-165.

Ballschmiter K (1992) Transport and fate of organic compounds in the global environment, *Angewandte Chemie - International Edition in English*, **31**, 487-664.

Ball A, Sedberry G, Zatcoff M, Chapman R, Carlin J (2000) Population structure of the wreckfish *Polyprion americanus* determined with microsatellite genetic markers, *Marine Biology*, **137**, 1077-1090.

Barrio Froján CR, Boyd SE, Cooper KM, Eggleton JD, Ware S (2008) Long-term benthic responses to sustained disturbance by aggreate extraction in an area off the east coast of the United Kingdom, *Estuarine, Coastal and Shelf Science*, **79**, 204-212.

Baumann PC (1998) Epizootics of cancer in fish associated with genotoxins in sediment and water, *Mutation Research*, **411**, 227-233.

Baumann PC, Harshbarger JC, Hartman KJ (1990) Relationship between liver tumors and age in brown bullhead populations from two lake Erie tributaries, *The Science of the Total Environment*, **94**, 71-97.

Beaumont AR (1991) Genetic studies of laboratory reared mussels, *Mytilus edulis*: heterozygote deficiencies, heterozygosity and growth, *Biological Journal of the Linnean Society*, **44**, 273-285.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans, *Molecular Ecology*, **13**, 969-980.

Beaumont WRC, Mann RHK (1984) The age, growth and diet of a freshwater population of the flounder, *Platichthys flesus* (L.), in southern England, *Journal of Fish Biology*, **25**, 607-616.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure, *Proceedings of the Royal Society of London: Series B*, **263**, 1619-1626.

Beebee T, Rowe G (2004) *An introduction to Molecular Ecology*, 1st edition, Oxford: Oxford University Press.

Beggs SE, Nash RDM (2007) Variability in settlement and recruitment of 0-group dab *Limanda limanda* L. in Port Erin Bay, Irish Sea, *Journal of Sea Research*, **58**, 90-99.

Behre K-E (2007) A new Holocene sea-level curve for the southern North Sea, *Boreas*, **36**, 82-102.

Bekhti M, Vianet R, Bouix G (1985) *Glugea stephani* (Hagenmüller, 1899), microsporidie parasite du flet *Platichthys flesus* (L., 1758) du littoral Languedocien. Importance du regime alimentaire de l'hôte dans le cycle saisonnier du parasite, *Vie et milieu*, **35**, 107-114.

Bekkevold D, André C, Dahlgren TG, Clausen LA, Torstensen E, Mosegaard H, Carvalho GR, Christensen TB, Norlinder E, Ruzzante DE (2005) Environmental correlates of population differentiation in Atlantic herring, *Evolution*, **59**, 2656-2668.

Belfiore N, Anderson S (2001) Effects of contaminants on genetic patterns in aquatic organisms: a review, *Mutation Research*, **489**, 97-122.

Belkhir K, Borsa P, Chikki L, Raufaste N, Bonhomme F (1996-2004) *Genetix 4.05, logiciel sous Windows TM pour la genetique des populations*, Montpellier: Laboratoire Genome, Populations, Interactions, CNRS UMR 5000, Universite de Montpellier II.

Bembo D, Carvalho G, Cingolani N, Arneri E, Giannetti G, Pitcher T (1996) Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters, *Marine Biology*, **126**, 529-538.

BEQUALM (2009), [Online], Available: http://www.bequalm.org/.

Berckmoes V, Scheirs J, Jordaens K, Blust R, Backeljau T, Verhagen R (2005) Effects of environmental pollution on microsatellite DNA diversity in wood mouse (*Apodemus sylvaticus*) populations, *Environmental Toxicology and Chemistry*, **24**, 2898-2907.

Bernal-Ramírez J, Adcock G, Hauser L, Carvalho G, Smith P (2003) Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents, *Marine Biology*, **142**, 567-574.

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data, *Molecular Biology and Evolution*, **15**, 1298-1311.

Bhisey R, Kotekar A, Buch S (2003) Polymorphism at CYP and GST gene loci and susceptibility to tobacco related cancers, *Proceedings of the Indian National Science Academy*, **B69**, 35-48.

Bickham JW, Sandhu S, N HPD, Chikki L, Arhwal R (2000) Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology, *Mutation Research*, **463**, 33-51.

Bilgin R (2007) Kgtests: a simple Excel macro program to detect signatures of population expansion using microsatellites, *Molecular Ecology Notes*, **7**, 416-417.

Blanco G, Presa P, Vázquez E, Sánchez J (1998) Allozyme heterozygosity and development in Atlantic salmon, *Salmo salar*, *Fish Physiology and Biochemistry*, **19**, 163-169.

Blanchet S, Rey O, Berthier P, Lek S, Loot G (2009) Evidence of parasite-mediated disruptive selection on genetic diversity in a wild fish population, *Molecular Ecology*, **18**, 1112-1123.

Blanquer A, Alayse J-P, Berrada-Rkhami O, Berrebi P (1992) Allozyme variation in turbot (*Psetta maxima*) and brill (*Scophthalmus rhombus*)(Osteichthyes, Pleuronectoformes, Scophthlamidae) throughout their range in Europe, *Journal of Fish Biology*, **41**, 725-736.

Bohonak AJ (2002) IBD (Isolation by Distance): a program for analyses of isolation by distance, *Journal of Heredity*, **93**, 153-154.

Bolle L, Bapper R, Witte J, van der Veer H (1994) Nursery grounds of dab (*Limanda limanda* L) in the sourthern North Sea, *Nertherlands Journal of Sea Research*, **32**, 299-307.

Borell Y, Pineda H, McCarthy I, Vázquez E, Sánchez J, Lizana G (2004) Correlations between fitness and heterozygosity at allozyme and microsatellite loci in the Atlantic salmon, *Salmo salar* L. , *Heredity*, **82**, 585-593.

Borsa P, Blanquer A, Berrebi P (1997) Genetic structure of flounders *Platichthys flesus* and *P. stellatus* at different geographic scales, *Marine Biology*, **129**, 233-246.

Bouck A, Vision T (2007) The molecular ecologists' guide to expressed sequence tags, *Molecular Ecology*, **16**, 907-924.

Bourret V, Couture P, Campbell P, Bernatchez L (2008) Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient, *Aquatic Toxicology*, **86**, 76-90.

Bouza C, Sánchez L, Martínez P (1997) Gene diversity analysis in natural populations and cultured stocks of turbot (*Scophthalmus maximus* L.), *Animal Genetics*, **28**, 28-36.

Braber L, de Groot S (1973) The food of live flatfish species (Pleuronectiformes) in the southern North Sea, *Netherlands Journal of Sea Research*, **6**, 163-172.

Bradbury IR, Campana SE, Bentzen P (2008) Estimating contemporary early life-history dispersal in an estuarine fish: integrating molecular and otolith elemental aproaches, *Molecular Ecology*, **17**, 1438-1450.

Brandström M, Ellegren H (2009) Genome-wide analysis of microsatellite polymorphims in chicken circumventing the ascertainment bias, *Genome Research*, **18**, 881-887.

Breen G, Shaw P, Li T, Shaw D, Collier D, Clair D (1999) Accuracy and sensitivity of DNA pooling with microsatellite repeats using capillary electrophoresis, *Molecular and Cellular Probes*, **13**, 359-365.

Breine J, Maes J, Quataert P, van den Bergh E, Simoiesn I, van Thuyne G, Belpaire C (2007) A fish-based assessment tool for the ecological quality of the brackish Schelde estuary in Flanders (Belgium), *Hydrobiologia*, **575**, 141-159.

Broeg K, Zander S, Diamant A, Körting W, Krüner G, Paperna I, Westernhagen H (1999) The use of fish metabolic, pathological and parasitological indices in pollution monitoring, *Helgoland Marine Research*, **53**, 171-194.

Brown J, Carrillo L, Fernand L, Horsburgh K, Hill A, Young E, Medler K (2003) Observations of the physical structure and seasonal jet-like circulation of the Celtic Sea and St. George's Channel of the Irish Sea, *Continental Shelf Research*, **23**, 533-561.

Brown MCL, Guttman S, Glenn TC (2001) Development and use of microsatellite DNA loci for genetic ecotoxicological studies of the fathead minnow (*Pimephales promela*), *Ecotoxicology*, **10**, 233-238.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history, *Trends in Ecology and Evolution*, **18**, 249-256.

Bryan G (1977) The pathogenesis of experimental bladder cancer, *Cancer Research*, **37**, 2813-2816.

Bucke D, Vethaak A, Lang T, Mellergaard S (1996) Common diseases and parasites of fish in the North Atlantic: training guide for identification. , in *ICES Techniques in Marine Environmental Sciences No. 19*, Copenhagen: ICES.

Buckley LJ, Smigielski AS, Halavik TA, Caldarone EM, Burns BR, Laurence GC (1991) Winter flounder, *Pseudopleuronectes americanus* reproductive success. II. Effects of spawning time and female size on size, composition and viability of eggs and larvae, *Marine Ecology Progress Series*, **74**, 125-135.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers, *Molecular Ecology Notes*, **5**, 684-687.

Camacho-Ibar V, McEvoy J (1996) Total PCBs in Liverpool Bay Sediments, *Marine Environmental Research*, **41**, 241-263.

Campos WLM, Kloppmann M, von Westernhagen H (1994) Inferences form the horizontal distribution of dab, *Limanda limanda* (L.) and flounder, *Platichthys flesus* (L.) larvae in the south-eastern North Sea, *Netherlands Journal of Sea Research*, **32**, 277-286.

Canino M, Spies I, Hauser L (2005) Development and characterization of novel di- and tetranucleotide microsatellite markers in Pacific Cod (*Gadus macrocephalus*), *Molecular Ecology Notes*, **5**, 908-910.

Carajaville MP, Hauser L, Carvalho GR, Hylland KOI, Lawrence AJ, Lowe D, Goksøyr A (2003) Genetic damage and the molecular-cellular response to pollution, in Lawrence A,Hemingway K (ed.) *Effects of pollution on fish. Molecular effects and population responses*, 1st edition.

Cardinale M, Doering-Arjes P, Kastowsky M, Mosegaard H (2004) Effects of sex, stock, and environment on the shape of known-age Atlantic cod (*Gadus morhua*) otoliths, *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 158-167.

Carreras-Carbonell J, Macpherson E, Pascual M (2006) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci, *Molecular Ecology*, **15**, 3527-3539.

Carvalho GR (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation, *Journal of Fish Biology*, **43**, 53-73.

Carvalho GR, Hauser L (1994a) Molecular genetics and the stock concept in fisheries, *Reviews in Fish Biology and Fisheries*, **4**, 326-350.

Carvalho G, Hauser L (1994b) Molecular genetics and the stock concept in fisheries, *Reviews in Fish Biology and Fisheries*, **4**, 326-350.

Carvalho GR, Hauser L (1998a) Advances in the molecular analysis of fish population structure, *Italian Journal of Zoology*, **65**, 21-33.

Carvalho G, Hauser L (1998b) Advances in the molecular analysis of fish population structure, *Italian Journal of Zoology*, **65**, 21-33.

Carvalho GR, Pitchet TJ (1995) *Molecular Genetics in Fisheries*, 1st edition, London: Chapman & Hall.

Casas L, Martínez P, Sánchez L (2005) Characterization of microsatellite markers derived from sequence databases for the European flounder (*Platichthys flesus*), *Molecular Ecology Notes*, **5**, 664-665.

Castillo AGF, Martinez JL, Garcia-Vazquez E (2004) Fine spatial structure of Atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci, *Marine Biotechnology*, **6**, 299-306.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures, *Evolution*, **21**, 550-570.

CEFAS (2000) *Monitoring and surveillance of non-radioactive contaminants in the aquatic environment and activities regulating the disposal of wastes at sea, 1997. Aquatic Environment Monitoring Report No. 52*, Lowestoft: CEFAS.

CEFAS (2003a) Fish Pathology and Disease Biomarkers 1999-2000, in *Monitoring of the quality of the marine environment, 1999-2000. Aquatic Environment Monitoring Report No. 54*, Lowestoft: CEFAS.

CEFAS (2003b) Measurements of DNA adducts in marine flatfish from offshore and estuarine locations, in *Monitoring of quality of the marine environment, 1999-2000. Aquatic Environment Monitoring Report No. 54*, Lowestoft: CEFAS.

CEFAS (2003c) The use of Ethoxyresorufin-O-deethylase (EROD) in fish as a biological effects monitoring tool, in *Monitoring of the quality of the marine environment. Science Series, Aquatic Environment Monitoring Report No. 54*, Lowestoft: CEFAS.

CEFAS (2005) Monitoring of the quality of the marine environment, in *Science Series, Aquatic Environment Monitoring Report No. 57*, Lowestoft: CEFAS.

Clarke KR, Warwick RM (2001) *Change in marine communities: an approach to statistical analysis and interpretation*, 2nd edition, Plymouth: Primer-E.

Coltman DW, Bowen WD, Wright JM (1998) Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites, *Proceedings of the Royal Society of London Series B*, **265**, 803-809.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population, *Evolution*, **53**, 1259-1267.

Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation, *Journal of Fish Biology*, **69**, 21-47.

Coombs AJ, Letcher HB, Nislow HK (2008) CREATE: a software to create input files from diploid genotypic data for 52 genetic software programs, *Molecular Ecology Resources*, **8**, 578-580.

Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure, *Bioinformatics*, **20**, 2363-2369.

Cornuet JM, Luikart G (1996) Description and power analysis of two test for detecting recent population bottlenecks from allele frequency data, *Genetics*, **144**, 2001-2014.

Corsolini S, Ademollo N, Romeo T, Greco S, Focardi S (2005) Persistent organic pollutants in edible fish: a human and environmental health problem, *Microchemical Journal*, **79**, 115-123.

Costa M, Cabral H, Drake P, Economou A, Fernández-Delgado C, Gordo L, Marchand J, Thiel R (2002) Recruitment and production of commercial species in estuaries, in Elliot M,Hemingway K (ed.) *Fishes in Estuaries*, Blackwell Science Ltd.

Coughlan J, McCarthy E, McGregor D, O'Donoghue P, Galvin P, Fitzgerald R, Daemen E, Imsland A, Stefansson S, Cross T (1996) Four polymorphic microsatellites in turbot *Scophthalmus maximus*, *Animal Genetics*, **27**, 441.

Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, Guinness FE, Clutton-Brock TH (1998) Microsatellites reveal heterosis in red deer, *Proceedings of the Royal Society of London Series B*, **265**, 489-495.

Coyer JA, Hoarau G, Beszteri B, Pearson G, Olsen JL (2009) Expressed sequence tag-derived polymorphic SSR markers for *Fucus serratus* and amplificantion in other species of *Fucus*, *Molecular Ecology Resources*, **9**, 168-170.

Crawford N (2009) SMoG-D: Software for the measurement of Genetic Diversity, *Molecular Ecology Resources*, submitted.

Cronin MA, Bickham JW (1998) A population genetic analysis on the potential for a crude oil spill to induce heritable mutations and impact natural populations, *Ecotoxicology*, **7**, 259-278.

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology, *Annual Review of Genomics and Human Genetics*, **1**, 281-328.

Chessel D, Dufour A-B, Thioulouse J (2004) The ade4 package - I: One-table methods, *R-News*, **4**, 5-10.

Chevolot M, Hoarau G, Rijnsdorp A, Stam W, Olsen J (2006) Phylogeography and population structure of thornback rays (*Raja clavata* L., Rajidae), *Molecular Ecology*, **15**, 3693-3705.

Chikako K, Kouichi Y, Koichi T, Yoichi N (2005) NQO1, MPO, and the risk of lung cancer: a HuGE review, *Genetics in Medicine*, **7**, 463-478.

Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics, *Aquaculture*, **255**, 1-29.

Daan N, Bromley PJ, Hislop JRG, Nielsen NA (1990) Ecology of North Sea fish, *Netherlands Journal of Sea Research*, **26**, 343-386.

D'Amato ME (2006) Demographic expansion and subtle differentiation in the long-tailed hake *Macruronus magellanicus*: evidence from microsatellite data, *Marine Biotechnology*, **8**, 189-201.

Danzmann RG, Ferguson MM, Allendorf FW (1988) Heterozygosity and component of fitness in a strain of rainbow trout, *Biological Journal of the Linnean Society*, **33**, 285-304.

David P (1998) heterozygosity-fitness correlations: new perspectives on old problems, *Heredity*, **80**, 531-537.

David P, Dauphin-Villemand C, Mesneau A, Meyran J (2003) Molecular approach to aquatic environmental bioreporting: differential response to environmental inducers of cytochrome P450 monooxygenase genes in the detritivorous subalpine planktonic crustacea, *Daphnia pulex*, *Molecular Ecology*, **12**, 2473-2481.

Davoodi F, Claireaux G (2007) Effects of exposure to petroleum hydrocarbons upon the metabolism of the common sole *Solea solea*, *Marine Pollution Bulletin*, **54**, 928-934.

De Innocentiis S, Sola L, Cataudella S, Bentzens P (2001) Allozyme and microsatellite loci provide discordant estimates of population differentiation in the endangered dusky grouper (*Epinephelus marginatus*) within the Mediterranean Sea, *Molecular Ecology*, **10**, 2163-3175.

de Veen JF (1978) On selective tidal transport in the migration of North Sea plaice (*Pleuronectes platessa*) and other flatfish species, *Netherlands Journal of Sea Research*, **12**, 115-147.

DEFRA (2002) *Safeguarding our seas*, London: Defra.

DEFRA (2009) *Marine and Coastal Access Bill Policy Document*, Defra.

Deniel C (1990) Comparative study of growth of flatfishes on the west coast of Brittany, *Journal of Fish Biology*, **37**, 149-166.

Depledge MH (1990) New approaches in ecotoxicology: can inter-individual physiological variability be used as a tool to investigate pollution effects?, *Ambio*, **19**, 251-252.

Depledge MH (1996) Genetic ecotoxicology: an overview, *Journal of Experimental Marine Biology and Ecology*, **200**, 57-66.

Desprez M (2000) Physical and biological impact of marine aggregate extraction along the Frech coast of the Eastern English Channel: short- and long-term post-dredging restoration, *ICES Journal of Marine Science*, **57**, 1428-1438.

Dethlefsen V, von Westernhagen H, Cameron P (1996) Malformations in North Sea pelagic fish embryos during the period 1984-1995, *ICES Journal of Marine Science*, **53**, 1024-1035.

DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals, *Journal of Fish Biology*, **56**, 461-473.

DeWoody JA, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations, *Molecular Ecology Notes*, **6**, 951-957.

Di S, Knowles JF (1994) The karyotype of the dab (*Limanda limanda* L.), *Journal of Fish Biology*, **44**, 354-356.

Dimsoski P, Toth GP (2001) Development of DNA-based microsatellite marker technology for studies of genetic diversity in stressor impacted populations, *Ecotoxicology*, **10**, 229-232.

Distelman-Menachem T, Shapira T, Laitman Y, Kaufman B, Barak F, Tavtigian S, Friedman E (2009) Analysis of *BRCA1/BRCA2* genes' contribution to breats cancer susceptibility in high risk Jewish Ashkenazi women, *Familial Cancer*, **8**, 127-133.

Dixon TJ, Taggart JB, George SG (2002) Application of real time PCR determination to assess interanimal variabilities in CYP1A induction in the European flounder (*Platichthys flesus*), *Marine Environmental Research*, **54**, 267-270.

Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*, *Molecular Biology and Evolution*, **18**, 1353-1364.

D'Surney SJ, Shugart LR, Theodorakis CW (2000) Genetic markers and genotyping methodologies: an overview, *Ecotoxicology*, **10**, 201-204.

Du Corbier FA, Stentiford GD, Lyons BP, Rotchell JM (2005) Isolation of the retinoblastoma cDNA from the marine flatfish dab (*Limanda limanda*) and evidence of mutational alterations in liver tumours, *Environmental Science and Technology*, **39**, 9785-9790.

Duftner N, Sefc K, Koblmüller S, Nevado B, Verheyen E, Phiri H, Sturmbauer C (2006) Distinct population structure in a phenotypically homogeneous rock-dwelling cichlid fish from lake Tanganyika, *Molecular Ecology*, **15**, 2381-2395.

Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations, *Molecular Biology and Evolution*, **18**, 672-675.

Eggens M, Bergman A, Vethaak D, van der Weiden M, Celander M, Boon J (1995) Cytochrome P4501A indices as biomarkers of contaminant exposure: results of a field study with plaice (*Pleuronectes platessa*) and flounder (*Platichthys flesus*) from the southern North Sea, *Aquatic Toxicology*, **32**, 221-225.

Ellegren H (2002a) Microsatellite evolution: a battle between replication slippage and point mutation, *Trends in Genetics*, **18**, 70.

Ellegren H (2002b) Mismatch repair and mutational bias in microsatellite DNA, *Trends in Genetics*, **18**, 552.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses, *Heredity*, **99**, 125-132.

England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential, *Conservation Genetics*, **4**, 595-604.

Espiñeira M, González-Lavín N, Vieites JM, Santaclara FJ (2008) Development of a method for the genetic identification of flatfish species on the basis of mitochondrial DNA sequences, *Journal of Agricultural and food chemistry*, **56**, 8954-8961.

Estoup A, Jarne P, Cornuet JM (2002a) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis, *Molecular Ecology*, **11**, 1591-1604.

EU Commission (2001) Organic contaminants in sewage sludge for agricultural use.

European Parliament (2008) Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Stategy Framework Directive), *Official Journal of the European Union*, **L 164**, 19-40.

Evans IA, Jones RS, Mainwaring-Burton R (1972) Passage of bracken fern toxicity into milk, *Nature*, **237**, 107-108.

Evans IA, Mason J (1965) Carcinogenic activity of bracken, *Nature*, **208**, 913-914.

Everaarts JM (1995) DNA integrity as a biomarker of marine pollution: strand breaks in seastar (*Asterias rubens*) and dab (*Limanda limanda*), *Marine Pollution Bulletin*, **31**, 431-438.

Exadactylos A, Geffen A, Thorpe J (1998) Population structure of the Dover sole, *Solea solea* L., in a background of high gene flow, *Journal of Sea Research*, **40**, 117-129.

Exadactylos A, Thorpe JP (2001) Allozyme variation and genetic inter-relationships between seven flatfish species (Pleuronectiformes), *Zoological Journal of the Linnean Society*, **132**, 487-499.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide, *Nature Reviews*, **7**, 745-758.

Excoffier L, Smouse PE, Quattro JM (1992a) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data, *Genetics*, **131**, 479-491.

Excoffier L, Smouse P, Quattro JM (1992b) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data, *Genetics*, **131**, 479-491.

Eyre-Walker A (2006) The genomic rate of adaptive evolution, *Trends in Ecology and Evolution*, **21**, 569-575.

FAO (2008) Marine genetic resources in areas beyond national jurisdiction as realted to marine biodiversity and the susteinable use of living marine resources, in *The state of World Fisheries and Aquaculture 2008*, Rome: Food and Agriculture Organization of the United Nations.

Fasching PA, Gayther S, Pearce L, Schildkraut JM, Goode E, Thiel F, Chenevix-Trench G, Chang-Claude J, Wang-Gohrke S, Ramus S, Pharoah P, Berchuck A (2009) Role of genetic polymorphisms and ovarian cancer susceptibility, *Molecular Oncology*, **3**, 171-181.

Feist SW, Lang T, Stentiford GD, Köhler A (2004) Biological effects of contaminants: Use of liver pathology of the European flatfish dab (*Limanda limanda* L.) and flounder (*Platichthys flesus* L.) for monitoring, in *ICES Techniques in Marine Environmental Sciences No. 38*, Copenhagen: ICES.

Feist SW, Lyons BP, Bignell J, Stentiford GD (2008) Fish health status in the North Sea and Irish Sea 2006, in *Monitoring of the quality of the marine environment, 2005-2006. Science Series. Aquatic Environment Monitoring Report No. 60*, Lowestoft: CEFAS.

Feist SW, Stentiford GD (2005) Fish pathology and disease biomarkers, in *Monitoring of the quality of the marine environment, 2002-2003. Science Series. Aquatic Environment Monitoring Report No. 57*, Lowestoft: CEFAS.

Felsenstein J (1989) PHYLIP-Phylogeny inference package (version 3.2), *Cladistics*, **5**, 164-166.

Florin A-B, Höglund J (2007) Absence of population structure of turbot (*Psetta maxima*) in the Baltic Sea, *Molecular Ecology*, **16**, 115-126.

Florin A-B, Höglund J (2008) Population structure of flounder (*Platichthys flesus*) in the Baltic Sea: differences among demersal and pelagic spawners, *Heredity*, **101**, 27-38.

Förlin L, Celander M (1993) Induction of cytochrome P450 1A in teleosts: environmental monitoring in swedish fresh, brackish and marine waters, *Aquatic Toxicology*, **26**, 41-56.

Foss A, Imsland AK, Nævdal G (1998) Population genetic studies of the Atlantic halibut in the North Atlantic Ocean, *Journal of Fish Biology*, **53**, 901-905.

Fox CJ, Dickey-Collas M, Winpenny AJ (1997) Spring plankton surveys of the Irish Sea in 1995: the distribution of fish eggs and larvae, *Science Series Technical Report. CEFAS, Lowestoft*, **104**, 106pp.

Franckowiak RP, Sloss BL, Bozek MA, Newman SP (2009) Temporal effective size estimates of a managed walleye *Sander vitreus* population and implications for genetic-based management, *Journal of Fish Biology*, **74**, 1086-1103.

Franklin A (2005) The concentration of mercury in fish taken from Liverpool Bay, UK in 2002. , in *Monitoring of the quality of the marine environment, 2002-2003. Aquatic Environment Monitoring Report No. 57*, Lowestoft: CEFAS.

Fraser D, Bernatchez L (2001) Adaptive evolutionary conservation: Towards a united concept for defining conservation units, *Molecular Ecology*, **10**, 2741-2752.

Frasier TR (2008) STORM: software for testing hypothesis of relatedness and mating patterns, *Molecular Ecology Resources*, **8**, 1263-1266.

Fratini S, Zane L, Ragionieri L, Vannini M, Cannicci S (2008) Relationship between heavy metal accumulation and genetic variability decrease in the intertidal crab *Pachygrapsus marmoratus* (Decapoda; Grapsidae), *Estuarine, Coastal and Shelf Science*, **79**, 679-686.

Funes V, Zuasti E, Catanese G, Infante C, Manchado M (2004) Isolation and characterization of ten microsatellite loci for Senegal sole (*Solea senegalensis* Kaup), *Molecular Ecology Notes*, **4**, 339-341.

Gagnon MM, Holdway DA (2000) EROD induction and biliary metabolite excretion following exposure to the water accommodated fraction of crude oil and to chemically dispersed crude oil, *Archives of Environmental Contamination and Toxicology*, **38**, 70-77.

Galarza JA, Carreras-Carbonell J, Macpherson E, Pascual M, Roques S, Turner GF, Rico C (2009) The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species, *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1473-1478.

Galpin O, Whitaker CJ, Whitaker R, Kassab JY (1990) Gastric-cancer in Gwynedd - possible links with bracken, *British Journal of Cancer*, **61**, 737-740.

Galvin P, Sadusky T, McGregor D, Cross T (1995) Population genetics of Atlantic cod using amplified single locus minisatellite VNTR analysis, *Journal of Fish Biology*, **47**, 200-208.

Galleguillos RA, Ward RD (1982) Genetic and morphological divergence between populations of the flatfish *Platichthys flesus* (L.) (Pleuronectidae), *Biological Journal of the Linnean Society*, **17**, 395-408.

Garcia de Leaniz C, Fleming I, Einum S, Verspoor E, Jordan W, Consuegra S, Aubin-Horth N, Lajus D, Letcher B, Youngson A, Webb J, Vøllestad L, Villanueva B, Ferguson A, Quinn T (2007)

A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation, *Biological Reviews*, **82**, 173-211.

Garcia-Vazquez E, Izquierdo J, Perez J (2006) Genetic variation at ribosomal genes supports the existence of two different European subspecies in the megrim *Lepidorhombus whiffiagonis* , *Journal of Sea Research*, **56**, 59-64.

Gardeström J, Dahl U, Kotsalainen O, Maxson A, Elfwing T, Grahn M, Bengtsson B-E, Breitholz M (2008) Evidence of pollution genetic effects of long-term exposure to contaminated sediments - A multi-endpoint study with copepods, *Aquatic Toxicology*, **86**, 426-436.

Gardeström J, Gorokhova E, Gilek M, Grahn M, Bengtsson B-E, Breitholtz M (2006) A multilevel approach to predict toxicity in copepod populations: assessment of growth, genetics, and population structure, *Aquatic Toxicology*, **79**, 41-48.

Garoia F, Marzola S, Guarniero I, Trentini M, Tinti F (2005) Isolation of polymorphic DNA microsatellites in the common sole *Solea vulgaris*, *Molecular Ecology Notes*, **6**, 144-146.

Geffen AJ, Jarvis K, Thorpe JP, Leah RT, Nash RDM (2003) Spatial differences in the trace element concentrations of Irish Sea plaice, *Pleuronectes platessa*, and whiting, *Merlangius merlangus*, otoliths, *Journal of Sea Research*, **50**, 245-254.

George S, Gubbins M, MacIntosh A, Reynolds W, Sabine V, Scott A, Thain J (2004) A comparison of pollutant biomarker responses with transcriptional responses in European flounders (*Platichthys flesus*) subjected to estuarine pollution, *Marine Environmental Research*, **58**, 571-575.

George S, Leaver M (2002) Allelic variations in the plaice UGT1B1 gene, *Marine Environmental Research*, **54**, 259-262.

George SG, Taylor B (2002) Molecular evidence for multiple UDP-glucoronosyltransferase gene families in fish, *Marine Environmetal Research*, **54**, 253-257.

Giæver M, Stien J (1998) Population genetic substructure in blue whiting based on allozyme data, *Journal of Fish Biology*, **52**, 272-295.

Gibson RN (2005) *Flatfishes: Biology and Exploitation*, 1st edition, New Dehli: Blackwell Publishing.

Gibson RN, Robb ML, Wennhage H, Burrows MT (2002) Ontogenetic changes in depth distribution of juvenile flatfishes in relation to predation risk and temperature on a shallow-water nursery ground, *Marine Ecology Progress Series*, **229**, 233-244.

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits, *Science*, **298**, 2345-2349.

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference, *Journal of Heredity*, **88**, 335-342.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci, *Genetics*, **139**, 463-471.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995b) Genetic absolute dating based on microsatellites and teh origin of modern humans, *Proccedings of the National Academy of Sciences of the United States of America*, **92**, 6723-6727.

Gosling E, Astanei I, Was A (2008) Genetic variability in Irish populations of the invasive zebra mussel, *Dreissena polymorpha*: discordant estimates of population differentiation from allozymes and microsatellites, *Freshwater Biology*, **53**, 1303-1315.

Goudet J (1995) FSTAT (vers. 1.2): a computer program to calculate F-statistics, *Journal of Heredity*, **86**, 485-486.

Goudet J (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*, [Online], Available: http://www.unil.ch/izea/softwares/fstat.html.

Goudet J, Keller L (2002) The correlation between inbreeding and fitness: does allele size matter?, *Trends in Ecology and Evolution*, **17**, 201-202.

Guillot G (2008) Inference of structure in subdivided populations at low level of genetic differentiation. The correlated allele frequencies model revisited., *Bioinformatics*, **24**, 2222-2228.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics, *Genetics*, **170**, 1261-1280.

Guillot G, Mortier F, Estoup A (2005b) Geneland: A computer package for landscape genetics, *Molecular Ecology Notes*, **5**, 708-711.

Guillot G, Santos F, Estoup A (2008) Analysisng georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface, *Bioinformatics*, **24**, 1406-1407.

Guinand B, Lemaire C, Bonhomme F (2004) How to detect polymorphisms undergoing selection in marine fishes? A review of methods and case studies, including flatfishes, *Journal of Sea Research*, **51**, 167-182.

Guinand B, Rolland JL, Bonhomme F (2008) Genetic structure of the common sole (*Solea solea*) in the Bay of Biscay: Nurseries as units of selection?, *Estuarine Coastal and Shelf Science*, **78**, 316-326.

Guo D-C, Milewicz D (2003) Methodology for using a universal primer to label amplified DNA segments for molecular analysis, *Biotechnology Letters*, **25**, 2079-2083.

Gyllensten U (1985) The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous, and freshwater species, *Journal of Fish Biology*, **26**, 691-699.

Gysels ES, Hellemans B, Pampoulie C, Volckaert FAM (2004) Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the colonization of the Mediterranean and the North Sea, *Molecular Ecology*, **13**, 403-417.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), *Nucleic Acids Symposium Series*, **41**, 95-98.

Hansen MM, Fraser DJ, Als TD, Mensberg KLD (2008) Reproductive isolation, evolutionary distinctiveness an setting conservation priorities: the case of the European lake whitefish and the endangered North Sea houting (*Coregonus* spp.), *BMC Evolutionary Biology*, **8**, Artcl. 137.

Hansen MM, Hemmer-Hansen J (2007) Landscape genetics goes to sea, *Journal of Biology*, **6**, Artcl. 6.

Hauser L, Adcock G, Smith P, Bernal-Ramírez J, Carvalho G (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*), *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11742-11747.

Hauser L, Carvalho G (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts, *Fish and Fisheries*, **41**, 333-362.

Hauser L, Seamons T, Dauer M, Naish K, Quinn T (2006) An empirical verification of population assignment methods by marking and parentage data: hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA, *Molecular Ecology*, **15**, 3157-3173.

Hawkins AJS, Rusin J, Bayne BL, Day AJ (1989) The metabolic/physiological basis of genotype-dependent mortality during copper exposure in *Mytilus edulis*, *Marine Environmental Research*, **23**, 253-257.

Hawkins WE, Walker WW, Overstreet RM, Lytle TF, Lytle JS (1988) Dose-related carcinogenic effects of water-borne benzo[a]pyere on livers of two small fish species, *Ecotoxicology and Environmental Safety*, **16**, 219-231.

Haynes D, Johnson J (2000) Organochlorine, heavy metal and polyaromatic hydrocarbon pollutant concentrations in the Great Barrier Reef (Australia) environment: a review, *Marine Pollution Bulletin*, **41**, 267-278.

Hedrick PW (1999a) *Genetics of Populations*, 2nd edition, Sudburry: Jones and Bartletts Publishers.

Hedrick PW (1999b) Perspective: Highly variable loci and their interpretation in evolution and conservation, *Evolution*, **53**, 313-318.

Hedrick PW (2005) A standarized genetic differentiation measure, *Evolution*, **59**, 1633-1638.

Heller R, Siegismund H (2009) Relationship between three measures of genetic differentiation Gst, Dest, and G'st: how wrong have we been?, *Molecular Ecology*, **18**, 2080-2083.

Hellou J, Mackay D, Fowler B (1995) Bioconcentration of polycyclic aromatic compounds from sediments to muscle of finfish, *Environmental Science and Technology*, **29**, 255-2560.

Hellou J, Payne JF, Upshall C, Fancey LL, Hamilton C (1994) Bioaccumulation of aromatic hydrocarbons from sediments: a dose-response study with flounder (*Pseudopleuronectes americanus*), *Archives of Environmental Contamination and Toxicology*, **27**, 477-485.

Hellou J, Warren WG (1997) Polycyclic aromatic compounds and saturated hydrocarbons in tissues of flatfish: insights on environmental exposure, *Marine Environmental Research*, **43**, 11-25.

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Leschcke V (2007a) Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.), *Heredity*, **99**, 592-600.

Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V (2007b) Evolutionary mechanisms shaping the genetic populations structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.), *Molecular Ecology*, **16**, 3104-3118.

Henderson P (1998) On the variation in dab *Limanda limanda* recruitment: a zoogeographic study, *Journal of Sea Research*, **40**, 131-142.

Hill AE, Brown J, Fernand L (1996) The western Irish Sea gyre: a retention system for Norway lobster (*Nephrops norvegicus*), *Oceanologica Acta*, **19**, 357-368.

Hill AE, Brown J, Fernand L (1997) The summer gyre in the western Irish Sea: Shelf sea paradigms and management implications, *Estuarine Coastal and Shelf Science*, **44**, 83-95.

Hinten GN, Hale MC, Gratten J, Mossman JA, Lowder BV, Mann MK, Slate J (2007) SNP-SCALE: SNP scoring by colour and length exclusion, *Molecular Ecology Notes*, **7**, 377-388.

Hinz H, Kröncke I, Ehrich S (2005) The feeding strategy of dab *Limanda limanda* in the southern North Sea: Linking stomach contents to prey availability in the environment, *Journal of Fish Biology*, **67**, 125-145.

Hitchcock DR, Bell S (2004) Physical impacts of marine aggregate dredging on seabed resources in coastal deposits, *Journal of Coastal Research*, **20**, 101-104.

Hoarau G, Boon E, Jongma DN, Ferber S, Palsson J, van der Veer HW, Rijnsdorp AD, Stam WT, Olsen JL (2005) Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.), *Proceedings of the Royal Society of London: Series B*, **272**, 497-503.

Hoarau G, Cook D, Stam WT, Olsen JL (2002a) New microsatellite primers for plaice, *Pleuronectes platessa* L. (Teleostei: Pleuronectidae), *Molecular Ecology Notes*, **2**, 60-61.

Hoarau G, Coyer JA, Veldsink JH, Stam, T W, Olsen JL (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*, *Molecular Ecology*, **16**, 3606-3616.

Hoarau G, Piquet AMT, van der Veer HW, Rijnsdorp AD, Stam WT, Olsen JL (2004) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: a comparison of resolving

power between microsatellites and mitochondrial DNA data, *Journal of Sea Research*, **51**, 183-190.

Hoarau G, Rijnsdorp AD, van der Veer HW, Stam WT, Olsen JL (2002b) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites reveal large-scale spatial and temporal homogeneity, *Molecular Ecology*, **11**, 1165-1176.

Htun-Han M (1978) The reproductive biology of the dab *Limanda limanda* (L.) in the North Sea: seasonal changes in the ovary, *Journal of Fish Biology*, **13**, 351-359.

Hudson-Edwards K, Macklin MG, Curtis CD, Vaughan DJ (1996) Processes of formation and distribution of Pb-, Zn-, Cd-, and Cu-bearing minerals in the Tyne Basin, Northeast England: Implications for metal-contaminated river systems, *Environmental Science and Technology*, **30**, 72-80.

Hughes TP, Bellwood DR, Folke C, Steneck RS, Wilson J (2005) New paradigms for supporting the resilience of marine ecosystems, *Trends in Ecology and Evolution*, **20**, 380-386.

Hutchinson WF, Carvalho GR, Rogers SI (2001) Marked genetic structuring in localised spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites, *Marine Ecology Progress Series*, **223**, 251-260.

Hutchinson WF, van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*), *Proceedings of the Royal Society of London Series B*, **270**, 2125-2132.

Islam MS, Tanaka M (2004) Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: a review and systhesis, *Marine Pollution Bulletin*, **48**, 624-649.

Iyengar A, Piyapattanakorn S, Heipel D, Stone D, Howell B, Child A, Maclean N (2000a) A suite of highly polymorphic microsatellite markers in turbot (*Scophthalmus maximus* L.) with potential for use across several flatfish species, *Molecular Ecology*, **9**, 368-371.

Iyengar A, Piyapattanakorn S, Stone D, Heipel D, Howell B, Baynes S, Maclean N (2000b) Identification of microsatellite repeats in Turbot (*Scophthalmus maximus*) and Dover Sole (*Solea solea*) using a RAPD-based technique: Characterization of microsatellite markers in Dover Sole, *Marine Biotechnology*, **2**, 49-56.

Jackson JBC, Kirby MX, Berger WH, Bjondal KA, Botsford LW, Bourque BJ, Bradbury RH, Cooke R, Erlandson J, Esters JA, Hughes TP, Kidwell S, Lange CB, Lenihan HS, Pandolfi JM, Peterson CH, Steneck RS, Tegner MJ, Warner RR (2001) Historical overfishing and the recent collapse of coastal ecosystems, *Science*, **293**, 629-638.

Jager Z (1998) Accumulation of flounder larvae (*Platichthys flesus* L.) in the Dollard (Ems estuary, Wadden Sea), *Journal of Sea Research*, **43**, 43-57.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back, *Trends in Ecology and Evolution*, **11**, 424-429.

Järup L (2003) Hazards of heavy metal contamination, *British Medical Bulletin*, **68**, 167-182.

Johansson H, Surget-Groba Y, Gow JL, Thorpe RS (2008) Development of microsatellite markers in the St Lucia anole, *Anolis luciae*, *Molecular Ecology Resources*, **8**, 1408-1410.

Johnson FM (1998) The genetic effects of environmental lead, *Mutation Research/Reviews in Mutation Research*, **410**, 123-140.

Johnson LL, Landahl JT, Kubin LA, Horness BH, Myers MS, Collier TK, Stein JE (1998) Assessing the effects of anthropogenic stressors on Puget Sound flatfish populations, *Journal of Sea Research*, **39**, 125-137.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers, *Bioinformatics*, **24**, 1403-1405.

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method, *Heredity*, **101**, 92-103.

Jombart T, Pontier D, Dufour A-B (2009) Genetic markers in the playground of multivariate analysis, *Heredity*, **102**, 330-341.

Jones PD (2006) Water quality and fisheries in the Mersey estuary, England: a historical perspective, *Marine Pollution Bulletin*, **53**, 144-154.

Jonsson G (1966) Contribution to the biology of the dab in Icelandic waters, *Rit Fiskideildar*, **IV**, 3-36.

Jorde PE, Knutsen H, Espeland SH, Stenseth NC (2007) Spatial scale of genetic structuring in coastal cod *Gadus morhua* and geographic extent of local populations, *Marine Ecology Progress Series*, **343**, 229-237.

Jørgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V (2005) Marine landscape and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea, *Molecular Ecology*, **14**, 3219-3234.

Jost L (2008) Gst and its relatives do not measure differentiation, *Molecular Ecology*, **17**, 4015-4026.

Jost L (2009) *D* vs. *Gst*: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009), *Molecular Ecology*, **18**, 2088-2091.

Judd N, Griffith W, Faustman E (2004) Contribution of PCB exposure from fish consumption to total dioxin-like dietary exposure, *Regulatory Toxicology and Pharmacology*, **40**, 125-135.

Kaiser MJ, Ramsay K (1997) Opportunistic feeding by dabs within areas of trawl disturbance: possible implications for increased survival, *Marine Ecology Progress Series*, **152**, 307-310.

Kaiser MJ, Spencer BE (1995) Survival of by-catch from a beam trawl, *Marine Ecology Progress Series*, **126**, 31-38.

Kalantzi OI, Alcock RE, Johnston PA, Santillo D, Stringer RL, Thomas GO, Jones KC (2001) The global distribution of PCBs and organochlorine pesticides in butter, *Environmental Science and Technology*, **35**, 1013-1018.

Kalendar R (2009) *FastPCR software for PCR primer and probe design and repeat search*, [Online], Available: http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm.

Kalinowski ST (2002a) Evolutionary and statistical properties of three genetic distances, *Molecular Ecology*, **11**, 1263-1273.

Kalinowski ST (2002b) How many alleles per locus should be used to estimate genetic distances?, *Heredity*, **88**, 62-65.

Kaplinski L, Andreson R, Puurand T, Remm M (2005) MultiPLX: automatic grouping and evaluation of PCR primers, *Bioinformatics*, **21**, 1701-1702.

Kenchington E, Harding G, Jones M, Prodöhl P (2009) Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*, *Molecular Ecology*, **18**, 1654-1667.

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population, *Proceedings of the National Academy of Sciences of Unites States of America*, **75**, 2868-2872.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation, *Molecular Ecology*, **10**, 807-821.

Kirby MF (2003) The use of Ethoxyresorufin-O-deethylase (EROD) in fish as a biological effects monitoring tool, in *Monitorin of the quality of the marine environment, 2000-2001. Science Series. Aquatic Environment Monitoring Report No. 56*, Lowestoft: CEFAS.

Kirby MF, Matthiessen P, Neall P, Tylor T, Allchin CR, Kelly CA, Maxwell DL, Thain JE (1999) Hepatic EROD activity in flounder (*Platichthys flesus*) as an indicator of contaminant exposure in English estuaries, *Marine Pollution Bulletin*, **38**, 676-686.

Kirby MF, Smith AJ, Barry J, Katsiadaki I, Lyons BP, Scott AP (2006) Differential sensitivity of flounder (*Platichthys flesus*) in response to oestrogenic chemical exposure: an issue for design and interpretation of monitoring and research programmes, *Marine Environmental Research*, **62**, 315-325.

Kite-Powell HL, Fleming LE, Backer LC, Faustman EM, Hoagland P, Tsuchiya A, Younglove LR, Wilcox BA, Gast RJ (2008) Linking the oceans to public health: current efforts and future directions, *Environmental Health*, **7**, S6.

Knutsen H, Jorde P, Albert O, Hoelzel A, STenseth N (2007) Population genetic structure in the North Atlantic Greenland halibut (*Reinhardtius hippoglossoides*): influenced by oceanic current systems?, *Canadian Journal of Fisheries and Aquatic Sciences*, **64**, 857-866.

Knutsen H, Jorde P, André C, Stenseth C (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod, *Molecular Ecology*, **12**, 385-394.

Koehler A (2004) The gender-specific risk to liver toxicity and cancer of flounder (*Platichthys flesus* L.) at the German Wadden Sea coast, *Aquatic Toxicology*, **70**, 257-276.

Köhler A, Deisemann H, Lauritzen B (1992) Histological and cytochemical indices of toxic injury in the liver of dab *Limanda limanda*, *Marine Ecology Progress Series*, **91**, 141-153.

Kondraganti S, Kathirvel M, Jiang W, Barrios R, Moorthy B (2005) Effects of 3-Methylcholanthrene on gene expression profiling in the rat using cDNA microarrays analyses, *Chemical Research in Toxicology*, **18**, 1634-1641.

Kotoulas G, Bonhomme F, Borsa P (1995) Genetic structure of the common sole *Solea vulgaris* at different geographic scales, *Marine Biology*, **122**, 361-375.

LaHood ES, Moran P, J O, Grant S, Park LK (2002) Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results, *Molecular Ecology Notes*, **2**, 187-190.

Laloë D, Jombart T, Dufour A-B, Moazami-Goudarzi K (2007) Consensus genetic structuring and typological value of markers using multiple co-inertia analysis, *Genetics Selection Evolution*, **39**, 545-567.

Lallias D, Stockdale R, Boudry P, Beaumont A, Lapègue S (2009) Characterization of 27 microsatellite loci in the European flat oyster *Ostrea edulis*, *Molecular Ecology Resources*, **9**, 960-963.

Langston W, Burt G, Pope N (1999) Bioavailability of metals in sediments fo the Dogger Bank (Central North Sea): a mesocosm study, *Estuarine, Coastal and Shelf Science*, **48**, 519-540.

Larno V, Laroche J, Launey S, Flammarion P, Devaux A (2001) Responses of chub (*Leuciscus cephalus*) populations to chemical stresss, assessed by genetic markers, DNA damage and cytochrome P4501A Induction, *Ecotoxicology*, **10**, 145-158.

Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK, Kruhøffer M, Grønkjær P, George SG, Dyrskjøt L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*), *Molecular Ecology*, **16**, 4678-4683.

Larsen PF, Nielsen EE, Williams TD, Loeschcke V (2008) Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*), *Heredity*, **101**, 247-259.

Larsson LC, Laikre D, Palm S, André C, Carvalho GR, Ryman N (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus, *Molecular Ecology*, **16**, 1135-1147.

Laslett RE (1995) Concentrations of dissolved and suspended particulate Cd, Cu, Mn, Ni, Pb and Zn in surface waters around the coasts of England and Wales and in Adjacent Seas, *Estuarine, Coastal and Shelf Science*, **40**, 67-85.

Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y (2002) Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism, *American Genetic Association*, **93**, 40-47.

Lawrence AJ, Elliott M (2003) Introduction and conceptual model, in Lawrence A (ed.) *Effects of pollution on fish*, 1st edition, Blackwell Science Ltd.

Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a "continuous" population under isolation by distance., *Molecular Biology and Evolution*, **20**, 491-502.

Lemaire C, Allegrucci G, Naciri M, Bahri-Sfar L, Kara H, Bonhomme F (2000) Do discrepancies between microsatellites and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)?, *Molecular Ecology*, **9**, 457-467.

Leonardi M, Tarifeno E, Vera J (2009) Diseases of the Chilean flounder, *Paralichthys adspersus* (Steindachner, 1867), as a biomarker of marine coastal pollution near the Itata River (Chile): part II. Histopathological lesions, *Archives of Environmental Contamination and Toxicology*, **56**, 546-556.

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within Genes: Structure, Function, and Evolution, *Molecular Biology and Evolution*, **21**, 991-1007.

Li YC, Korol AB, Fazima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, *Molecular Ecology*, **11**, 2453-2465.

Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat *Sprattus sprattus*: differentiation across a steep environmental gradient in a small pelagic fish, *Marine Ecology Progress Series*, **379**, 213-224.

Liu ZJ, Cordes JF (2004) DNA marker technology and their application in aquaculture genetics, *Aquaculture*, **238**, 1-37.

Lønning S, Kjørsvik E, Falk-Petersen I (1988) A comparative study of pelagic and demersal eggs from common marine fishes in northern Norway, *Sarsia*, **73**, 49-60.

Lugon-Moulin N, Brünner H, Wyttenbach A, Hausser J, Goudet J (1999) Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora: Soricidae), *Molecular Ecology*, **8**, 419-431.

Luikart G, Cornuet J-M (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data, *Conservation Biology*, **12**, 228-237.

Lundy CJ, Moran P, Rico C, Milner RS, Hewitt GM (1999) Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish, *Molecular Ecology*, **8**, 1889-1898.

Lynch HT, Casey MJ, Snyder CL, Bewtra C, Lynch JF, Butts M, Godwin AK (2009) Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management, *Molecular Oncology*, **3**, 97-137.

Lynch HT, Smyrk T (1996) Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome), *Cancer*, **78**, 1149-1167.

Lyons BP, Stentiford GD, Bignell J, Sivier DB, Devlin MJ, Lowe D, Beesley A, Pascoe CK, Moore MNGE (2006) A biological effects monitoring survey of Cardigan Bay using flatfish histopathology, cellular biomarkers and sediment bioassays: findings of the Prince Madog Prize 2003, *Marine Environmental Research*, **62**, S342-S346.

Lyons BP, Stewart C, Kirby MF (2000) 32P-postlabelling analysis of DNA adducts and EROD induction as biomarkers of genotoxin exposure in dab (*Limanda limanda*) from British coastal waters, *Marine Environmental Research*, **50**, 575-579.

Maes GE, Raeymaekers JAM, Pampoulie C, Seynaeve A, Goemans G, Belpaire C, Volckaert FAM (2005) The catadromous European eel *Anguilla anguilla* (L.) as a model for freshwater evolutionary ecotoxicology: relationship between heavy metal bioaccumulation, condition and genetic variability, *Aquatic Toxicology*, **73**, 99-114.

MAFF (1987) Irish Sea status report of the Marine Pollution Monitoring Management Group, *Aquatic Environment Monitoring Report*, **17**, 84.

MAFF (1990) *Monitoring and surveillance of non-radioactive contaminants in the aquatic environment, 1984-1987. Aquatic Environment Monitoring Report No. 22*, Lowestoft: MAFF.

MAFF (1995) *Monitoring for diseases in marine and freshwater fish, 1992. Aquatic Environment Monitoring Report No. 46*, Lowestoft: MAFF Directorate of Fisheries Research.

Maggs CA, Castilho R, Foltz D, Henzler C, Taimour Jolly M, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa, *Ecology*, **89**, S108-S122.

Magoulas A, Tsimenides N, Zouros E (1996) Mitochondrial DNA phylogeny aand the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*), *Molecular Biology and Evolution*, **13**, 178-190.

Mainguy J, Côté D, Coltman D (2009) Multilocus heterozygosity, parental relatedness and individual fitness components in a wild mountain goat, *Oreamnos americanus* population, *Molecular Ecology*, **18**, 2297-2306.

Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignement tests and multilocus genotypes, *Conservation Biology*, **16**, 650-659.

Manel S, Schwartz M, Luikart G, Taberlet P (2003) Lanscape genetics: combining landscape ecology and populations genetics, *Trends in Ecology and Evolution*, **18**, 189-197.

Manni F, Guérard E (2004) *Barrier vs. 2.2. Manual of the user.*, Paris: Population genetic team, Museum of Mankind (Musée de l'Homme).

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm", *Human Biology*, **76**, 173-190.

Marchand J, Codling I, Drake P, Elliot M, Pihl L, Rebelo J (2002) Environmental quality of Estuaries, in Elliot M,Hemingway K (ed.) *Fishes in Estuaries*, Blackwell Science Ltd.

Marchand J, Tanguy A, Charrier G, Quiniou L, Plee-Gauthier E, Laroche J (2006) Molecular identification and expression of differentially regulated genes on the European flounder, *Platichthys flesus*, submitted to pesticide exposure, *Marine Biotechnology*, **8**, 275-294.

Marchand J, Tanguy A, Laroche J, Quiniou L, Moraga D (2003) Responses of European flounder *Platichthys flesus* populations to contamination in different estuaries along the Atlantic coast of France, *Marine Ecology Progress Series*, **260**, 273-284.

Mariani S, Hutchinson WF, Hatfield EMC, Ruzzante DE, Simmonds EJ, Dahlgren TG, André C, Brigham J, Torstensen W, Carvalho GR (2005) North Sea herring population structure revealed by microsatellite analysis, *Marine Ecology Progress Series*, **303**, 245-257.

Martinez-Lara E, Leaver M, George S (2002) Evidence from heterologous expression of glutathione S-transferases A and A1 of the plaice (*Pleuronectes platessa*) that their

endogenous role is in detoxification of lipid peroxidation products, *Marine Environmental Research*, **54**, 263-266.

Mattiangeli V, Bourke E, Ryan A, Mork J, Cross T (2000) Allozyme analyses of the genus *Trisopterus*: taxonomic status and population structure of the poor cod, *Journal of Fish Biology*, **56**, 474-494.

Mattiangeli V, Galvin P, Ryan A, Mork J, Cross T (2002) VNTR variability in Atlantic poor cod (*Trisopterus minutus minutus*) throughout its range: single locus minisatellite data suggest reproductive isolation for the Faroe Bank population, *Fisheries Research*, **58**, 185-191.

McCairns RJS, Bernatchez L (2008) Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large scale estuarine environment, *Molecular Ecology*, **17**, 3901-3916.

McGowan C, Reith M (1999) Polymorphic microsatellite markers for Atlantic halibut, *Hippoglossus hippoglossus*, *Molecular Ecology*, **8**, 1761-1763.

Medina RF, Barbosa P, Christman M, Battisti A (2006) Number of individuals and molecular markers to use in genetic differentiation studies, *Molecular Ecology Notes*, **6**, 1010-1013.

Meirmans PG (2006) Using the AMOVA framework to estimate a standarized genetic differentiation measure, *Evolution*, **60**, 2399-2402.

Meirmans PG, van Tiernderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms, *Molecular Ecology Notes*, **4**, 792-794.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances, *Trends in Plant Science*, **12**, 1360-1385.

Meyer JN, Nacci DE, Di Giulio RT (2002) Cytochrome P4501A (CYP1A) in killifish (*Fundulus heteroclitus*): heritability of altered expression and relationship to survival in contaminated sediments, *Toxicological Sciences*, **68**, 69-81.

Michalakis Y, Excoffier L (1996) A genetic estimation of population subdivision using distances between alleles with special reference for microsatellite loci, *Genetics*, **142**, 1061-1064.

Miller JA, Banks MA, Gomez-Uchida D, Shanks AL (2005) A comparison of population structure in black rockfish (*Sebastes melanops*) as determined with otolith microchemistry and microsatellite DNA, *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 2189-2198.

Millner R, Walsh S, Díaz de Astarloa J (2005) Atlantic Flatfish Fisheries, in Gibson R (ed.) *Flatfishes: Biology and Exploitation*, 1st edition, New Dehli: Blackwell Publishing.

Missiaggia A, Grattapaglia D (2006) Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers, *Genetics and Molecular Research*, **5**, 283-291.

Mix MC (1986) Cancerous diseases in aquatic animals and their association with environmental pollutants: a critical literature review, *Marine Environmental Research*, **20**, 1-141.

Moore JW, Moore IA (1976) The basis of food selection in flounders, *Platichthys flesus* (L.), in the Severn Estuary, *Journal of Fish Biology*, **9**, 139-156.

Munkacsi AB, Stoxen S, May G (2008) *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas, *Proceedings of the Royal Society of London: Series B*, **275**, 1037-1046.

Myers MS, Anulacion BF, French BL, Reichert WL, Laetz CA, Buzitis J, Olson OP, Sol S, Collier TK (2008) Improved flatfish health following remediation of a PAH-contaminated site in Eagle Harbor, Washington, *Aquatic Toxicology*, **88**, 277-288.

Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities, *Nature*, **423**, 280-283.

Naish KA, Hard JJ (2008) Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes, *Fish and Fisheries*, **9**, 396-422.

Nakata H, Sakai Y, Miyawaki T, Takemura A (2003) Bioaccumulation and toxic potencies of polychlorinated biphenyls and polycyclic aromatic hydrocarbons in tidal flat and coastal ecosystems of the Ariake Sea, Japan, *Environmental Science and Technology*, **37**, 3513-3521.

Neall P, Rooke J, Kirby M (2003) The use of biomarkers in biological effects monitoring, in *Monitoring of the quality of the marine environment, 2000-2001; Science Series. Aquatic Environment Monitoring Report No. 56*, Lowestoft: CEFAS.

Nebert DW, McKinnon RA, Puga A (1996) Human drug-metabolizing enzyme polymorphisms: effects on risk of toxicity and cancer, *DNA and Cell Biology*, **15**, 273-280.

Neff BD, Fu P, Gross MR (2000) Microsatellite multiplexing in fish, *Transactions of the American Fisheries Society*, **129**, 584-593.

Nei M (1972) Genetic distance between populations, *American Naturalist*, **106**, 283-292.

Nei M (1973) Analysis of gene diversity in subdivided populations, *Proceeding of the National Academy of Sciences of the United States of America*, **70**, 3321-3323.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals, *Genetics*, **89**, 593-590.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities, *Annals of Human Genetics*, **47**, 253-259.

Neuparth M (2004) *A genetic ecotoxicological approach using crustacean and fish as model organisms*, Ph.D thesis edition, Lisboa, Portugal: Universidade Nova de Lisboa.

Nielsen EE, Hansen MM, Meldrup D (2006) Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms, *Molecular Ecology*, **15**, 3219-3229.

Nielsen EE, Nielsen PH, Meldrup D, Hansen MM (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea, *Molecular Ecology*, **13**, 585-595.

Nielsen EE, Wright PJ, Hemmer-Hansen J, Poulsen NA, Gibb IM, Meldrup D (2009) Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals, *Marine Ecology Progress Series*, **376**, 213-225.

Nikolic N, Féve K, Chevalet C, Høyheim B, Riquet J (2009) A set of 37 microsatellite DNA markers for genetic diversity and structure analysis of Atlantic salmon *Salmo salar* populations, *Journal of Fish Biology*, **74**, 458-466.

Nissling A, Westin L, Hjerne O (2002) Reproductive success in relation to salinity for three flatfish species, dab (*Limanda limanda*), plaice (*Pleuronectes platessa*), and flounder (*Platichthys flesus*), in the brackish water Baltic Sea, *Journal of Marine Science*, **59**, 93-108.

Nock NL, Cicek MS, Li L, Liu X, Rybicki BA, Moreira A, Plummer SJ, Casey G, Witte JS (2006) Polymorphisms in estrogen bioactivation, detoxification and oxidative DNA base excision repair genes and prostate cancer risk, *Carcinogenesis*, **27**, 1842-1848.

Nosil P, Funk D, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence, *Molecular Ecology*, **18**, 375-402.

Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats, *Evolution*, **59**, 705-719.

Nowak C, Vogt C, Pfenninger M, Schwenk K, Oehlmann J, Streit B, Oetken M (2009) Rapid genetic erosion in pollutant-exposed experimental chironomid populations, *Environmental Pollution*, **157**, 881-886.

Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers, *Genomics*, **30**, 450-458.

Ogden R (2008) Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry, *Fish and Fisheries*, **9**, 462-472.

Oliveira EJ, Gomes Pádua J, Zucchi MI, Vencovsky R, Carneiro Vieira ML (2006) Origin, evolution and genome distribution of microsatellites, *Genetics and Molecular Biology*, **29**, 294-307.

Olsen JB, Habichi C, Reynolds J, Seeb JE (2004) Moderately and highly polymorphic microsatellites provide discordant estimates of population divergence in sockeye salmon, *Oncorhynchus nerka*, *Environmental Biology of Fishes*, **69**, 261-273.

Olsen EM, Knutsen H, Gjosæter J, Jorde PE, Knutsen JA, Stenseth NC (2008) Small-scale biocomplexity in coastal Atlantic cod supporting a Darwinian perspective on fisheries management, *Evolutionary Applications*, **1**, 524-533.

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between Fst and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure, *Molecular Ecology*, **13**, 1799-1814.

Ortega-Salas A (1980) Seasonal changes in the common dab, *Limanda limanda* (L.) in Isle of Man waters, *Journal of Fish Biology*, **16**, 75-82.

OSPAR (2009) *JAMP guidelines for contaminant-specific biological effects (OSPAR agreement 2008-09)*, [Online], Available: http://www.ospar.org/.

OSPAR Commission (2000) *Quality Status Report*, London: OSPAR Commission.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears., *Molecular Ecology*, **4**, 347-354.

Paetkau D, Slade R, Burdens M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power, *Molecular Ecology*, **13**, 55-65.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations, *Genetics*, **147**, 1943-1957.

Pakeman RJ, Marrs RH, Howard DC, Barr CJ, Fuller RM (1996) The bracken problem in Great Britain: its present extent and future changes, *Applied Geography*, **16**, 65-86.

Pampoulie C, Stefánsson M, Jörundsdóttir T, Danilowicz B, Daníelsdóttir A (2008) Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L., *Biological Journal of the Linnean Society*, **94**, 315-329.

Papetti C, Susana E, Patarnello T, Zane L (2009) Spatial and temporal boundaries to gene flow between *Chaenocephalus aceratus* populations at South Orkney and South Shetlands, *Marine Ecology Progress Series*, **376**, 269-281.

Pardo GB, Casas L, Fortes GG, Bouza C, Martínez P, Clark MS, Sánchez L (2005) New microsatellite markers in turbot (*Scophthalmus maximus*) derived from an enriched genomic library and sequence databases, *Molecular Ecology Notes*, **5**, 62-64.

Pauly D, Christensen V, Dalsgaard J, Froese R, Torres FJ (1998) Fishing down marine food webs, *Science*, **279**, 860-863.

Pauly D, Watson R, Alder J (2005) Global trends in world fisheries: impacts on marine ecosystems and food security, *Philosophical Transactions of the Royal Society B*, **360**, 5-12.

Payseur BA, Jing P (2009) A genomewide comparison of population structure at STRPs and nearby SNPs in humans, *Molecular Biology and Evolution*, **26**, 1369-1377.

Peakall R, Smouse P (2001) *GenAlEx V5.04: Genetic Analysis in Excel. Population genetics software for teaching and research*, [Online], Available: http://www.anu.edu.au/BoZo/GenAlEx/.

Pearce NJG, Mann VL (2006) Trace metal variations in the shells of *Ensis siliqua* record pollution and environmental conditions in the sea to the west of mainland Britain, *Marine Pollution Bulletin*, **52**, 739-755.

Peles JD, Towler W, Guttman SI (2003) Population genetic structure of earthworm (*Lumbricus rubellus*) in soils contaminated by heavy metals, *Ecotoxicology*, **12**, 379-386.

Penrith ML, Bastianello SS, Penrith MJ (1993) Hepatic lipoidosis and fatty infiltration of organs in a captive African stonefish, *Synanceja verrucosa* Bloch & Schnider, *Journal of Fish Diseases*, **17**, 171-176.

Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J (1997) Microsatellite variation and the differentiation of modern humans, *Human Genetics*, **99**, 1-7.

Permuth-Wey J, Egan K (2009) Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis, *Familial cancer*, **8**, 109-117.

Perry AL, Low PJ, Ellis JR, Reynolds JD (2005) Climate change and distribution shifts in marine fishes, *Science*, **308**, 1912-1915.

Pharoah PDP, Day NE, Duffy S, Easton DF, Ponder BAJ (1997) Family history and the risk of breast cancer: a systematic review and meta-analysis, *International Journal of Cancer*, **71**, 800-809.

Phillips DJH, Segar DA (1986) Use of bio-indicators in monitoring conservative contaminants: Programme designs imperatives, *Marine Pollution Bulletin*, **17**, 10-17.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection, *Journal of Heredity*, **95**, 536-539.

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequecy data, *Journal of Heredity*, **90**, 502-503.

Pogson GH, Fevolden SE (1998) DNA heterozygosity and growth rate in the Atlantic cod *Gadus morhua* (L), *Evolution*, **52**, 915-920.

Pogson GH, Fevolden S-E (2003) Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation a the pantophysin (*Pan*I) locus, *Molecular Ecology*, **12**, 63-74.

Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic Cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci, *Genetics*, **139**, 375-385.

Potter DM, Baird MS (2000) Carcinogenic effects of ptaquiloside in bracken fern and related compounds, *British Journal of Cancer*, **83**, 914-920.

Poulsen NA, Nielsen EE, Schierup MH, Leschcke V, Grønkjær P (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*), *Molecular Ecology*, **15**, 321-331.

Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud, *Proceedings of the Royal Society of London: Series B*, **267**, 1699-1704.

Primmer CR, Møller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds, *Molecular Ecology*, **5**, 365-378.

Pritchard JK, Stephens M, Donolly P (2000) Inference of population structure using multilocus genotype data, *Genetics*, **155**, 945-959.

Pulsford AL, Crampe M, Langston A, Glynn PJ (1995) Modulatory effects of disease, stress, copper, TBT and vitamin E on the immune system of flatfish, *Fish & Shellfish Immunology*, **5**, 631-643.

R Development Core Team (2009) *R: a language and environment for statistical computing.*, URL: http://www.R-project.org edition, Vienna, Austria: R Foundation for Statistical Computing, Available: http://www.R-project.org.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes, *Procceding of the National Academy of Sciences of the United States of America*, **94**, 9197-9201.

Rasmussen LH, Kroghsbo S, Frisvad JC, Hansen HCB (2003) Occurence of the carcinogenic Bracken constituent ptaquiloside in fronds, topsoils and organic soil layers in Denmark, *Chemosphere*, **51**, 117-127.

Raufaste N, Bonhomme F (2000) Properties of bias and variance of two multiallelic estimators of *Fst*, *Theoretical Population Biology*, **57**, 285-296.

Raymond M, Rousset F (1995) An exact test for population differentiation, *Evolution*, **49**, 1280-1283.

Reichert WL, Myers MS, Peck-Miller K, French B, Anulacion BF, Collier TK, Stein JE, Varanasi U (1998) Molecular epizootiology of genotoxic events in marine fish: linking contaminant exposure, DNA damage, and tissue-level alterations, *Mutation Research*, **411**, 215-225.

Reich DE, Feldman MW, Goldstein DB (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions, *Molecular Biology and Evolution*, **16**, 453-466.

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 8119-8124.

Reid DP, Pongsomboon S, Jackson T, McGowan C, Murphy C, Martin-Robichaud D, Reith M (2005) Microsatellite analysis indicates an absence of population structure among *Hippoglossus hippoglossus* in the north-west Atlantic, *Journal of Fish Biology*, **67**, 570-576.

Renshaw MA, Saillant E, Broughton RE, Gold JR (2006) Application of hypervariable genetic markers to forensic identification of 'wild' from hatchery-raised red drum, *Sciaenops ocellatus*, *Forensic Science International*, **156**, 9-15.

Reusch RBH, Ehlers A, Hämmerli A, Worm B (2005) Ecosystem recovey after climatic extremes enhanced by genotypic diversity, *Proceedings of the Natural Academy of Sciences: Series B*, **102**, 2826-2831.

Rijnsdorp AD, Vethaak AD, van Leeuwen PI (1992) Population biology of dab *Limanda limanda* in the southeastern North Sea, *Marine Ecology Progress Series*, **91**, 19-35.

Rijnsdorp AD, Witthames RR (2005) Ecology of Reproduction, in Gibson R (ed.) *Flatfishes: Biology and Exploitation*, 1st edition, Blackwell Publishing.

Ritland K (2002) Estimation of gene frequency and heterozygosity from pooled samples, *Molecular Ecology Notes*, **2**, 370-372.

Roark SA, Andrews JF, Guttman SI (2001) Population genetics structure of the western mosquitofish, *Gambusia affinis*, in a highly channelized portion of the San Antonio River in San Antonio, TX., *Ecotoxicology*, **10**, 223-227.

Roman J, Palumbi S (2004) A global invader at home: population structure of the green crab, *Carcinus maenas*, in Europe, *Molecular Ecology*, **13**, 2891-2898.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry, *American Jounal of Human Genetics*, **73**, 1402-1422.

Rossetto M, McNally J, Henry R (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae, *Theoretical and Applied Genetics*, **104**, 61-66.

Rotchell JM, Du Corbier FA, Stentiford GD, Lyons BP, Liddle AR, Ostrander GK (2009) A novel population health approach: using fish *retinoblastoma* gene profiles as a surrogate for humans, *Comparative Biochemistry and Physiology, Part C*, **149**, 134-140.

Rotchell JM, Stagg RM, Craft JA (1995) Chemically-induced genetic damage in fish: isolation and characterization of the dab (*Limanda limanda*) Ras gene, *Marine Pollution Bulletin*, **31**, 457-459.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance, *Genetics*, **145**, 1219-1228.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux, *Molecular Ecology Resources*, **8**, 103-106.

Rowe G, Beebee T (2005) Intraspecific competition disadvantages inbred natterjack toad (*Bufo calamita*) genotypes over outbred ones in a shared pond environment, *Journal of Animal Ecology*, **74**, 71-76.

Rozen S, Skaletsky HJ (2000) PRIMER3 on the WWW for general users and for biologist programmers (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), in Krawetz S,Misener S (ed.) *Bioinformatic Methods and Protocols: Methods in Molecular Biology*, Totowa, New Jersey: Humana Press.

Rudin N, Inman K (2002) *An introduction to forensic DNA analysis*, 2nd edition, CRC Press.

Ruzzante DE, Mariani S, Bekkevold D, André C, Mosegaard H, Clausen LAW, Dahlgren TG, Hutchinson WF, Hatfield EMC, Torstensen E, Brigham J, Simmonds EJ, Laikre L, Larsson LC, Stet RJM, Ryman N, Carvalho G (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring, *Proceedings of the Royal Society of London: Series B*, **273**, 1459-1464.

Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence, *Fisheries Research*, **43**, 79-97.

Ryk C, Kumar R, Sanyal S, Berggren de Verdier P, Hemminki K, Larsson P, Steineck G, Hou S (2006) Influence of polymorphism in DNA repair and defence genes on *p53* mutations in bladder tumours, *Cancer Letters*, **241**, 142-149.

Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations, *Evolution*, **62**, 2250-2259.

Ryman N, Leimar O (2009) Gst is still a useful measure of genetic differentiation - a comment on Jost's *D*, *Molecular Ecology*, **18**, 2084-2087.

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation, *Molecular Ecology*, **6**, 600-602.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic Salmon (*Salmo salar*), *Journal of Heredity*, **98**, 692-704.

Saito K, Nagao T, Matoba M, Koyama K, Natori S, Murakami T, Saiki Y (1989) Chemical assay of ptaquiloside, the carcinogen of *Pteridium aquilinum*, and the distribution of related compounds in the Pteridaceae, *Phytochemistry*, **28**, 1605-1611.

Scott AP, Sanders M, Stentiford GD, Reese RA, Katsiadaki I (2007) Evidence for estrogenic endocrine disruption in an offshore flatfish, the dab (*Limanda limanda* L.), *Marine Environmental Research*, **64**, 128-148.

Scheringer M (2009) Long-range transport of organic chemicals in the environment, *Environmental Toxicology and Chemistry*, **28**, 667-690.

Schiedek D, Sundelin B, Readman J, Macdonald R (2007) Interactions between climate change and contaminants, *Marine Pollution Bulletin*, **54**, 1845-1856.

Schlacher T, Mondon J, Connolly R (2007) Estuarine fish health assessment: Evidence of wastewater impacts based on nitrogen isotopes and histopathology, *Marine Pollution Bulletin*, **54**, 1762-1776.

Schlezinger JJ, Blickarz CE, Mann KK, Stegeman JJ (2000) NF-kB (REL) homologues in scup (*Stenotomus chrysops*) and possible activation of NF-kB by AH receptor agonists, *Marine Environmental Research*, **50**, 61-81.

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps, *Genetics*, **160**, 753-763.

Schmidt V, Zander S, Körting W, Broeg K, von Westernhagen H, Dizer H, Hansen P, Skouras A, Steinhagen D (2003) Parasites of flounder (*Platichthys flesus* L.) from the German Bight, North Sea, and their potential use in biological effects monitoring, *Helgoland Marine Research*, **57**, 262-271.

Schmitt CJ, Wildhaber ML, Hunn JB, Nash T, Tieger MN, Steadman BL (1993) Biomonitoring of lead-contaminated Missouri Streams with an assay for Erythrocyte δ-aminolevulinic acid dehydratase activity in fish blood, *Archives of Environmental Contamination and Toxicology*, **25**, 464-475.

Schopen GCB, Bovenhuis H, Visker MHPW, van Arendonk JAM (2008) Comparison of information content for microsatellites and SNPs in poultry and cattle, *Animal Genetics*, **39**, 451-453.

Schoske R, Vallone P, Ruitberg C, Butler J (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci, *Analytical and Bioanalytical Chemistry*, **375**, 333-343.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments, *Nature Biotechnology*, **18**, 233-234.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management, *Treends in Ecology and Evolution*, **22**, 25-33.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results, *Conservation Genetics*, **10**, 441-452.

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population, *Molecular Ecology*, **14**, 503-511.

Seisay MBD (2001) *Population ecology of dab (Limanda limanda L.) in the Eatern Irish Sea, North Wales.*, Ph.D. Thesis edition, Bangor, UK: Univerisity of Wales, Bangor.

Selkoe KA, Henzler C, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations, *Fish and Fisheries*, **9**, 363-377.

Seyfert AL, Cristescu MEA, Frisse L, Schaack S, Thomas WK, Lynch M (2008) The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*, *Genetics*, **178**, 2113-2121.

Shaw PW, Turan C, Wright JM, O'Connell M, Carvalho GR (1999) Microsatellite DNA analysis of population struture in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses, *Heredity*, **83**, 490-499.

Sheader DL, Gensberg K, Lyons BP, Chipman JK (2004) Isolation of differentially expressed genes from contaminant exposed European flounder by suppressive, substractive hybridisation, *Marine Environmental Research*, **58**, 553-557.

Sheader DL, Williams TD, Lyons BP, Chipman JK (2006) Oxidative stress response of European flounder (*Platichthys flesus*) to cadmium determined by a custom cDNA microsarray, *Marine Environmental Research*, **62**, 33-44.

Shikano T, Chiyokubo T, Taniguchi N (2001) Temporal changes in allele frequency, genetic variation and inbreeding depression in small populations of the guppy, *Poecilia reticulata*, *Heredity*, **86**, 153-160.

Shikano T, Taniguchi N (2002) Relationships between genetic variation measured by microsatellite DNA markers and a fitness-related trait in the guppy (*Poecilia reticulata*), *Aquaculture*, **209**, 77-90.

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci, *Molecular Biology and Evolution*, **12**, 914-920.

Shriver M, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach, *Genetics* , **134**, 983-993.

Simán SE, Povey AC, Ward TH, Margison GP, Sheffield E (2000) Fern spore extracts can damage DNA, *British Journal of Cancer*, **83**, 69-73.

Simpson MG, Parry M, Kleinkauf A, Swarbreck D, Walker P, Leah RT (2000) Pathology of the liver, kidney and gonad of flounder (*Platichthys flesus*) from a UK estuary impacted by endocrine disrupting chemicals, *Marine Environmental Research*, **50**, 283-287.

Skouras A, Lang T, Vobach M, Danischewski D, Wosniok W, Scharsack J, Steinhagen D (2003) Assessment of some innate immune responses in dab (*Limanda limanda* L.) from the North Sea as part of an integrated biological effects monitoring, *Helgoland Marine Research*, **57**, 181-189.

Slatkin M (1987) Gene flow and the geographic structure of natural populations, *Science*, **236**, 787-792.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies, *Genetics*, **139**, 457-462.

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene, *Genetical Research*, **23**, 23-35.

Smith PJ, Robertson SG, Horn PL, Bull B, Anderson OF, Stanton BR, Oke CS (2002) Multiple techniques for determining stock relationships between orange roughy, *Hoplostethus atlanticus*, fisheries in the eatern Tasman Sea, *Fisheries Research*, **58**, 119-140.

Smith J, Shackley S (2006) Effects of the closure of a major sewage outfall on sublittoral soft sediment benthic communities, *Marine Pollution Bulletin*, **52**, 645-658.

Sotelo CG, Calo-Mata P, Chapela JM, Pérez-Martín RI, Rehbein H, Hold GL, Russell VJ, Pryde S, Quinteiro J, Izquierod M, Rey-Méndez M, Rosa C, Santos AT (2001) Identification of flatfish (Pleuronectiforme) species using DNA-based techniques, *Journal of Agriculture and Food Chemistry*, **49**, 4562-4569.

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks, *Molecular Ecology*, **9**, 1517-1528.

Srogi K (2007) Monitoring of environmental exposure to polycyclic aromatic hydrocarbons: a review, *Environmental Chemistry Letters*, **5**, 169-195.

Staton JL, Schizas NV, Chandler GT, Coull BC, Quattro JM (2001) Ecotoxicology and population genetics: The emergence of "Phylogeographic and Evolutionary Ecotoxicology", *Ecotoxicology*, **10**, 217-222.

Stein JE, Reichert WL, Nishimoto M, Varanasi U (1990) Overview of studies on liver carcinogenesis in English sole from Puget Sound; evidence for a xenobiotic chemical etiology II: biochemical studies, *The Science of the Total Environment*, **94**, 51-69.

Stentiford GD, Bignell JP, Lyons BP, Feist SW (2009) Site-specific disease profiles in fish and their use in environmental monitoring, *Marine Ecology Progress Series*, **381**, 1-15.

Stentiford GD, Longshaw M, Lyons B, Jones G, Green M, Feist SW (2003) Histopathological biomarkers in estuarine fish species for the assessment of biological effects of contaminants, *Marine Environmental Research*, **55**, 137-159.

Stenvik J, Wesmajervi M, Fjalestad K, Damsgård, Delghandi M (2006) Development of 25 gene-associated microsatellite markers of Atlantic cod (*Gadus morhua* L.), *Molecular Ecology Notes*, **6**, 1105-1107.

Stott D (2009) Environmental enforcement in the UK, *Journal of Environmental Monitoring*, **11**, 470-474.

Suk HY, Neff BD (2009a) Microsatellite genetic differentiation among populations of the Trinidadian guppy, *Heredity*, **102**, 425-434.

Svendäng H (2003) The inshore demersal fish community on the Swedish Skagerrak coast: regulation by recruitment from offshore sources, *Journal of Marine Science*, **60**, 23-31.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA, *Genetics*, **144**, 389-399.

Tanabe S (1988) PCB problems in the future: foresight from current knowledge, *Environmental Pollution*, **50**, 5-28.

Teh SJ, Adams SM, Hinton DE (1997) Histopathologic biomarkers in feral freshwater fish populations exposed to different types of contaminant stress, *Aquatic Toxicology*, **37**, 51-70.

Teschke M, Mukabayire O, Wiehe T, Tautz D (2008) Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans, *Genetics*, **180**, 1537-1545.

Theodorakis CW (2001) Integration of genotoxic and population genetic endpoints in biomonitoring and risk assessment, *Ecotoxicology*, **10**, 245-256.

Theodorakis CW, Blaylock BG, Shugart LR (1997) Genetic ecotoxicology I: DNA integrity and reproduction in mosquitofish exposed *in situ* to radionucleotides, *Ecotoxicology*, **6**, 205-218.

Theodorakis CW, Elbl T, Shugart LR (1999) Genetic Ecotoxicology IV: survival and DNA strand breakage is dependent on genotype in radionucleotide-exposed mosquitofish, *Aquatic Toxicology*, **45**, 279-291.

Theodorakis CW, Lee KL, Adams SM, Law CB (2006) Evidence of altered gene flow, mutation rate, and genetic diversity in redbreast sunfish from a Pulp-mill-contaminated river, *Environmental Science and Technology*, **40**, 337-386.

Theodorakis CW, Shugart LR (1997) Genetic Ecotoxicology II: population genetic structure in mosquitofish exposed *in situ* to radionucleotides, *Ecotoxicology*, **6**, 335-354.

Theodorakis CW, Shugart LR (1998) Genetic Ecotoxicology III: the relationship between DNA strand breaks and genotype in mosquitofish exposed to radiation, *Ecotoxicology*, **7**, 227-236.

Thomas M, Möller F, Wiehe T, Tautz D (2007) A pooling approach to detect signatures of selective sweeps in genome scans using microsatellites, *Molecular Ecology Notes*, **7**, 400-403.

Thorpe JP, Solé-Cava AM, Watts PC (2000) Exploited marine invertebrates: genetics and fisheries, *Hydrobiologia*, **420**, 165-184.

Thorrold SR, Latkoczy C, Swat PK, Jones CM (2001) Natal homing in a marine fish metapopulation, *Science*, **291**, 297-299.

Tsitrone A, Rousset F, David P (2001) Heterosis, marker mutational processes and population inbreeding history, *Genetics*, **159**, 1845-1859.

Tysklind N, Neuparth T, Ashcroft GR, Taylor MI, Lyons BP, McCarthy ID, Carvalho GR (2009a) Isolation and characterization of 28 new microsatellite markers for European flounder (*Platichthys flesus* L.), *Molecular Ecology Resources*, **9**, 1065-1068.

Tysklind N, Taylor MI, Lyons BP, McCarthy ID, Carvalho GR (2009b) Development of 30 microsatellite markers for dab (*Limanda limanda* L.): a key UK biomonitoring species, *Molecular Ecology Resources*, **9**, 951-955.

Untersee S, Pechenik JA (2007) Local adaptation and maternal effects in two species of marine gastropod (genus *Crepidula*) that differ in dispersal potetial, *Marine Ecology Progress Series*, **347**, 79-85.

Valdes A, Slatkin M, Freimer N (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited, *Genetics*, **133**, 737-749.

van der Land MA (1991) Distribution of flatfish eggs in the 1989 egg surveys in the southeastern North Sea, and mortality of plaice and sole eggs, *Netherlands Journal of Sea Research*, **27**, 277-286.

van der Molen J, Rogers S, Ellis J, Fox C, McCloghrie P (2007) Dispersal patterns of the eggs and larvae of spring-spawning fish in the Irish Sea, UK, *Journal of Sea Research*, **58**, 313-330.

van der Oost R, Beyer J, Vermeulen N (2003) Fish bioaccumulation and biomarkers in environmental risk assessment: a review, *Environmental Toxicology and Pharmacology*, **13**, 54-149.

van der Veer HW (1985) Impact of coelenterate predation on larval plaice *Pleuronectes platesa* and flounder *Platichthys flesus* stock in the western Wadden Sea, *Marine Ecology Progress Series*, **25**, 229-238.

van der Veer HW (1986) Immigration, settlement, and density-dependent mortality of a larval and early postlarval 0-group plaice (*Pleuronectes platesa*) population in the western Wadden Sea, *Marine Ecology Progress Series*, **29**, 223-236.

van der Veer HW, Berghahn R, Miller J, Rijnsdorp A (2000) Recruitment in flatfish, with special emphasis on North Atlantic species: Progress made by the Flatfish Symposia, *ICES Journal of Marine Science*, **57**, 202-215.

van der Veer HW, Bergman MJN, Dapper R, Witte JIJ (1991) Population dynamics of an intertidal 0-group flounder *Platichthys flesus* population in the western Dutch Wadden Sea, *Marine Ecology Progress Series*, **73**, 141-148.

van Herwerden L, Benzie J, Davies C (2003) Microsatellite variation and population genetic structure of the red throat emperor on the Great Barrier Reef , *Journal of Fish Biology*, **62**, 987-999.

van Oosterhout C, Hutchinson W, Wills D, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data, *Molecular Ecology Notes*, **4**, 535-538.

Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic Salmon (*Salmo salar* L.), *Molecular Biology and Evolution*, **22**, 1067-1076.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies, *Molecular Ecology*, **14**, 3623-3642.

Verneau O, Moreau C, Catzeflis F, Renaud F (1994) Phylogeny of flatfishes (Pleuronectiformes): comparisons and contradictions of molecular and morpho-anatomical data, *Journal of Fish Biology*, **45**, 685-696.

Verspoor E (1994) The evolution of genetic divergence at protein coding loci among anadromous and nonanadrous populations of Atlantic salmon *Salmo salar*, in Beaumont A (ed.) *Genetics and Evolution of Aquatic Organisms*, 1st edition, London: Chapman & Hall.

Vethaak AD, Jol JG, Pieters JP (2009) Long-term trends in the prevalence of cancer and other major diseases among flatfish in the southeastern North Sea as indicators of changing ecosystem health, *Environmental Science and Technology*, **43**, 2151-2158.

Viguera E, Canceill D, Ehrlich S (2001) Replication slippage involves DNA polymerase pausing and dissociation, *European Molecular Biology Organization Journal*, **20**, 2587-2595.

Vitale F, Börjesson P, Svedäng H, Casini M (2008) The spatial distribution of cod (*Gadus morhua* L.) spawning grounds in the Kattegat, eastern North Sea, *Fisheries Research*, **90**, 36-44.

von Westernhagen H, Dethlefsen V, Bade T, Wosniok W (2002) Species assemblages of pelagic fish embryos in the southern North Sea between 1984 and 2000., *Helgoland Marine Research*, **55**, 242-251.

Wagner CE, McCune AR (2009) Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes, *Evolution*, **63**, 1312-1326.

Waisberg M, Joseph P, Hale B, Beyersmann D (2003) Molecular and cellular mechanisms of cadmium carcinogenesis, *Toxicology*, **192**, 95-117.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species, *Journal of Heredity*, **9**, 438-450.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity, *Molecular Ecology*, **15**, 1419-1439.

Waples RS, Punt AE, Cope JM (2008) Integrating data into management of marine resources: how can we do it better?, *Fish and Fisheries*, **9**, 423-449.

Ward RD (2000) Genetics in fisheries management, *Hydrobiologia*, **420**, 191-201.

Ward DG, Wei G, Cheng Y, Billingham LJ, Martin A, Johnson P, Lyons B, Feist S, Stentiford G (2006) Plasma proteome analysis reveals the geographical origin and liver tumor status of dab (*Limanda limanda*) from UK marine waters, *Environmental Science and Technology*, **40**, 4031-4036.

Ward RD, Woodward M, Skibinski D (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes, *Journal of Fish Biology*, **44**, 213-232.

Watts PC, Nash RD, Kemp SJ (2001) Polymorphic microsatellite loci in the European plaice, *Pleuronectes platessa*, and their utility in flounder, lemon sole and Dover sole, *Journal of the Marine Biological Association of the United Kingdom*, **81**, 367-368.

Weetman D, Hauser L, Bayes MK, Ellis JR, Shaw PW (2006) Genetic population structure across a range of geographic scales in the commercially exploited marine gastropod *Buccinum undatum*, *Marine Ecology Progress Series*, **317**, 157-169.

Weilgart LS (2007) The impacts of anthopogenic ocean noise on cetateans and implications for management, *Canadian Journal of Zoology*, **85**, 1091-116.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure, *Evolution*, **38**, 1358-1370.

Weis JS, Weis P (1989) Tolerance and stress in a polluted environment: the case of the mummichog, *BioScience*, **39**, 89-95.

Whibley C, Pharoah P, Hollstein M (2009) p53 polymorphisms: cancer implications, *Nature Cancer Reviews*, **9**, 95-107.

Whitfield AK, Elliot M (2002) Fishes as indicators of environmental and ecological changes within estuaries: a review of progress and some suggestions for the future, *Journal of Fish Biology*, **61**, 229-250.

Wilson D, Donaldson L, Sepai O (1998) Should we be frightened of bracken? A review of the evidence, *Journal of Epidemiology and Community Health*, **52**, 812-817.

Willet K, Rogers J, Di Giulio R (2000) Differential phase I and phase II enzyme activities in brown bullhead and channel catfish exposed to β-naphthoflavone, *Marine Environmental Research*, **50**, 67-68.

Williams TD, Diab AM, George SG, Sabine V, Chipman JK (2007) Gene expression responses of European flounder (*Platichthys flesus*) to 17-β estradiol, *Toxicology Letters*, **168**, 236-248.

Wirgin I, Kreamer G, Garte S (1991) Genetic polymorphism of cytochrome *P450*IA in cancer-prone Hudson River tomcod, *Aquatic Toxicology*, **19**, 205-214.

Worm B, Barbier E, Beaumont N, Duffy J, Folke C, Halpern B, Jackson J, Lotze H, Micheli F, Palumbi S, Sala E, Selkoe K, Stchowicz J, Watson R (2006) Impacts of biodiversity loss on ocean ecosystem services, *Science*, **314**, 787-790.

Wright S (1943) Isolation by distance, *Genetics* , **28**, 114-138.

Wright S (1951) The genetical structure of populations, *Annals of Eugenics*, **15**, 323-354.

Wu RSS, Siu WHL, S SPK (2005) Induction, adaptation and recovery of biological responses: Implications for environmental monitoring, *Marine Pollution Bulletin*, **51**, 623-634.

Wu SH, Tsai SM, Hou MF, Lin HS, Hou LA, Ma H, Lin JT, Yeh FL, Tsai LY (2006) Interaction of genetic polymorphism in cytochrome P450 2E1 and glutathione S-Transferase M1 to breast cancer in Taiwanese women without smoking and drinking habits, *Breats Cancer Research and Treatment*, **100**, 93-98.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review, *Molecular Ecology*, **11**, 1-16.

Zulliger DE, Tanner S, Ruch M, Ribi G (2009) Genetic structure of the high dispersal Atlanto-Mediterranean sea star *Astropecten aranciacus* revealed by mitochondrial DNA sequences and microsatellite loci, *Marine Biology*, **156**, 597-610.